Applications of hidden Markov models in financial modelling

A thesis submitted for the degree of Doctor of Philosophy

by

Christina Erlwein

Department of Mathematical Sciences
School of Information Systems,
Computing & Mathematics
Brunel University

May 2008

To my parents

and

to Charles

Abstract

Various models driven by a hidden Markov chain in discrete or continuous time are developed to capture the stylised features of market variables whose levels or values constitute as the underliers of financial derivative contracts or investment portfolios. Since the parameters are switching regimes, the changes and developments in the economy as soon as they arise are readily reflected in these models. The change of probability measure technique and the EM algorithm are fundamental techniques utilised in the optimal parameter estimation. Recursive adaptive filters for the state of the Markov chain and other auxiliary processes related to the Markov chain are derived which in turn yield self-tuning dynamic financial models. A hidden Markov model (HMM)-based modelling set-up for commodity prices is developed and the predictability of the gold market under this setting is examined. An Ornstein-Uhlenbeck (OU) model with HMM parameters is proposed and under this set-up, we address two statistical inference issues: the sensitivity of the model to small changes in parameter estimates and the selection of the optimal number of states. The extended OU model is implemented on a data set of 30-day Canadian T-bill yields. An exponential of a Markov-switching OU process plus a compound Poisson process is put forward as a model for the evolution of electricity spot prices. Using a data set compiled by Nord Pool, we illustrate the vast improvements gained in incorporating regimes in the model. A multivariate HMM is employed as a framework in providing the solutions of two asset allocation problems; one involves the mean-variance utility function and the other entails the CVaR constraint. Finally, the valuation of credit default swaps highlights the important considerations necessitated by pricing in a regime-switching environment. Certain numerical schemes are applied to obtain approximations for the default probabilities and swap rates.

Contents

List of Tables

List of Figures

Acknowledgements

Several people made it possible for me to complete this thesis. First and foremost I would like to thank my supervisor Dr Rogemar Mamon for his guidance and research support, which had a profound impact on this work. His knowledge and insights helped me through various challenges in my research. This thesis has greatly benefited from his numerous excellent comments and suggestions. Furthermore I would like to thank him and the Department of Statistical and Actuarial Sciences at the University of Western Ontario, London, Canada for their hospitality during my research visit.

I would also like to thank Professor Fred Espen Benth for his valuable advice, many fruitful discussions and kind hospitality during my stay at the Centre of Mathematics for Applications, University of Oslo. The collaboration with him initiated new research interests and added new aspects to my work. I thank Dr Tak Kuen Siu with whom I had the opportunity to collaborate at the Department of Actuarial Mathematics and Statistics at Heriot-Watt University, Edinburgh for the hospitality, many helpful insights and stimulating discussions. I would like to acknowledge Professor Gautam Mitra, who has been very supportive to my research and gave me the opportunity to complete an industrial placement at OptiRisk Systems. Furthermore I would like to thank my examiners Dr Paresh Date and Professor Bernard Hanzon for their valuable comments and suggestions.

I gratefully acknowledge the financial support provided by a Brunel Research Initiative and Enterprise Fund (BRIEF) of Dr Rogemar Mamon and an E.U. Marie Curie Fellowship for Early Stage Researchers Training within NET-ACE (MEST-CT-2004-6724) at Brunel University under the coordination of Professor Geoff Rodgers.

I thank all my colleagues and friends at Brunel University, especially Katharina Schwaiger, Christian Valente, Luka Jalen, Leela Mitra, Nilgun Canakgoz and Rima Sheikh Rajab as well as Yeliz Yolcu Okur and Andrea Barth from University of Oslo, who all contributed to a very enjoyable working environment.

I would like to express my special thanks to my family and friends, especially to my parents Renate and Franz Erlwein, who constantly supported and encouraged me throughout all these years. Finally I am deeply grateful to Charles Cochran for his love, patience and unlimited support, that made the completion of this thesis possible.

Comments on works with other research collaborators

This thesis presents the product of research carried out during my PhD studies. I had the opportunity to work with researchers from various universities; these collaborations led to publications and technical reports, which are detailed below.

A modified version of chapter 3 appeared as publication [111]. Rogemar Mamon developed the filter equations and parameter estimates, whilst Bushan Gopulani contributed to the $ARMA$-model comparison and the Diebold-Kilian measure analysis. My contributions in the article are the verification and implementation of the HMM filters and DK-measures as well as the comparison with ARCH/ GARCH models.

Publication [61] is a modified version of chapter 5. Rogemar Mamon conceptualised the research plan, formulated the model and supervised and monitored the results. He also put forward major aspects of the statistical inference. My contributions in the article are the construction of the change of probability measures and the derivation of optimal parameter estimate equations as well as the implementation of the HMM filters. In addition, I calculated the formulae and numerical values for the Fisher information and the AIC.

Chapter 6 is an extended version of a technical report submitted for publication [62]. This research was carried out during a 3-month university placement at the University of Oslo in 2006. Fred Espen Benth specified the research plan and together with Rogemar Mamon supervised steps of the implementation and monitored the results. My contributions in the article are the formulation of the model framework, derivation of the optimal parameter estimate equations and the numerical implementation.

Research for the first part of chapter 7 was completed during my 3-month research visit in 2007 at the University of Western Ontario in collaboration with Rogemar Mamon and Matt Davison. The research problem was formulated by Rogemar Mamon and Matt Davison, they also monitored the numerical results. My contributions include the development of the algorithms for the filters that include multivariate series, implementation of the different investment strategies and the calculation of the optimal weight solution.

Parts of the second project in chapter 7 are a technical report [63], which emerged from an industry placement at OptiRisk. Gautam Mitra specified the research plan and monitored the results. Diana Roman contributed to the development of the CVaR framework. My contributions are the development and implementation of the scenario generator and the stochastic optimisation programs.

Major parts of chapter 8 constitute an article accepted for publication in NAAJ [139], which was initiated during my research visit at Heriot-Watt University, Edinburg in May and June 2006. Tak Kuen Siu derived the Esscher transform and supervised the calculation of numerical results. Rogemar Mamon formalised the theory and drafted the paper. My contributions in the article are the numerical solution and the implementation of the finite difference scheme and the Monte-Carlo method.

Nomenclature

The following notation will be used throughout the discussion of this work.

$\boldsymbol{A}^\top$	transpose of a matrix or a vector $\boldsymbol{A}$
$\mathbf{x}_k$	Markov chain
$k = 0, 1, ...$	discrete time
$\mathcal{M}$	state space of the Markov chain
χ	initial distribution of $\mathbf{x}$
h	number of forecast steps
$\mathbf{e}_i$	unit vector
$\mathbf{\Pi} = \{\pi_{ji}\}$	transition probability matrix
$\mathbf{v}_{k+1}$	martingale increment
N	number of states of the Markov chain
$(\Omega, \mathcal{F}, P)$	underlying probability space
$\mathcal{F}$	global filtration
$\mathcal{F}^y$	filtration generated by observation process
$\mathcal{F}^{\mathbf{x}}$	filtration generated by the Markov chain
$\mathcal{F}^W$	filtration generated by the Brownian motion$W$
$\bar{P}$	reference probability measure
$\{w_k\}$	sequence of IID standard normal random variables
$S(k)$	daily gold spot price
$\{y_k\}$	logarithmic returns of gold prices
$\boldsymbol{f}$	drift of $\{y_k\}$
$\boldsymbol{\sigma}$	volatility of $\{y_k\}$
Ξ_k	conditional expectation $\bar{\mathrm{E}}[\Lambda_k \mathbf{x}_k \vert \mathcal{F}_k^y]$
H	$\mathcal{F}-$adapted stochastic process
J_k^{sr}	process of number of jumps of a Markov chain $\mathbf{x}$
O_k^r	occupation time process of a Markov chain $\mathbf{x}$
$T_k^r(f)$	auxiliary process for a function f of an observation process
$\eta(H_k)$	unnormalised conditional expectation of H_k given $\mathcal{F}_k^y$ under $\bar{P}$, $\overline{\mathrm{E}}\big[H_k \bar{\Lambda}_k \mid \mathcal{F}_k^y\big]$
$\eta(\mathbf{x}_k)$	estimator of the state of the Markov chain
$\hat{H}_k$	conditional expectation $E[H_k \vert \mathcal{F}_k^y]$
r_t	continuous-time short term interest rate
a_t	rate of mean-reversion of interest rate process
δ_t	long-term mean of an interest rate process
u_t	volatility of an interest rate process
$\mathcal{R}_k$	complete filtration generated by r
$\mathcal{H}_k = \mathcal{F}_k^{\mathbf{x}} \vee \mathcal{R}_k$	global filtration generated by $\mathbf{x}$ and r

r_k	discretised short term interest rate
$\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\xi}$	parameters of discretised interest rate
$\rho 1$	set of model parameters for interest rate model
$SP(k)$	daily electricity spot price
$D(k)$	deterministic seasonal function
$Z(k)$	deseasonalised electricity log spot price
$d1_h, ..., d6_h$	parameters of sinusoidal function
s_h	factors in sinusoidal function
$\boldsymbol{\kappa}, \boldsymbol{\beta}, \boldsymbol{\zeta}$	parameters of the OU-process component for electricity spot prices
W	standard Brownian motion
Y_t	jump process
G_t	Poisson process
B	normally distributed jump size with mean μ_B and variance ζ_B^2
λ^P	intensity of a Poisson process
μ_Z	mean of observation process without jumps
ζ_z^2	variance of observation process without jumps
τ_m	random time of occurrence of the mth jump
b	denominator in reference probability measure $\bar{\Lambda}_k$
ρ_{elec}	set of model parameters $\alpha_i, \beta_i, \zeta_i^2, \mu_{B_i}, \zeta_{B_i}^2, \pi_{ji}$
ϕ	pdf of a standard normal distribution
$xp \in XP$	portfolio xp in portfolio set XP
$\text{loss}(xp, y)$	portfolio loss
tr	threshold in CVaR model
$b_\alpha(xp)$	$\alpha-$VaR
$c_\alpha(xp)$	$\alpha-$CVaR
$\breve{r}(t)$	instantaneous interest rate of a money market account
$\{V(t), t \geq 0\}$	firm's value process
$\mu(t), \varsigma(t)$	expected growth rate and volatility rate of firm's value
$U(t) := \ln(V(t)/V(0))$	continuously compounded return of firm's value V
$\mathbf{Q}$	generator matrix for continuous time Markov process
$\{\vartheta(t)\}_{t \in \mathcal{T}}$	real-valued stochastic process for Esscher transform
ϱ	first passage time of default
$SR = \frac{PL}{KL+ML}$	swap rate, with
PL	expected payoff from the contingent payment leg
KL	(sum of expected present values of the fee leg)$/SR$
ML	(sum of expected present values of the fee accruals)$/SR$
RR	constant rebate of underlying reference asset
L	exogenously given default barrier
$\mathcal{P}^\vartheta(\varrho \leq t_k \vert \mathcal{G}(t))$	conditional risk-neutral default probability
H_k	$I_{\{L \geq \min_{t \in [0,t_k]} V(t)\}}$, where I is the indicator function

Chapter 1

Introduction

A hidden Markov model (HMM) is a mathematical model in which the system being modelled contains a hidden Markov process. The parameters of the model are unknown and must be determined from a set of observable data. The technique has its origin in speech recognition and signal processing and has become more popular in mathematical finance in the last decade. Hamilton pioneered the application of HMMs in the analysis of economic time series in [79] and [80]. The main idea behind an HMM is that the latent state of the system and other unobservable information are hidden in an observation process, which is corrupted by some "noise". In the context of finance and economics, the observation process could be a univariate or multivariate financial time series. This hidden information is assumed to follow the dynamics of a finite-state Markov chain in discrete or continuous time. To extract and benefit from this information for making financial decisions, we need to filter the signals out of the observed data via an HMM filtering technique. Modelling in financial markets is mainly concerned with the determination of models, which are able to capture the empirically observed characteristics of financial time series. In the valuation of financial products, for example, the uncertainty of the underlying assets or variables needs to be taken into account. When measuring various types of financial risks and optimising portfolios, an accurate model for the underlying variables is also desired. Therefore, it is necessary that the model is able to reproduce the realistic features of the observed time series.

In this thesis, we demonstrate that HMMs can evidently capture these realistic

characteristics in several markets. The flexibility of an HMM can be attributed to its ability to allow regime shifts that may occur from time to time in the dynamics of the financial variables. Whilst HMM applications in finance have been explored by other researchers, this thesis proposes and develops further applications and extensions of certain HMM frameworks.

The motivation of the thesis in advancing HMM-modulated models in finance is supported by empirical studies in recent years. These empirical papers provide overwhelming evidence that incorporating regime shifts in modelling frameworks adds more capability in replicating the stylised behaviours of the data series. Driffill, et al. [45], amongst others conducted a study and found an increased number of realistic data features by employing a regime-switching model for interest rates. In a regime-switching model set-up, parameters are able to take different values depending on the state of an economy. Generally, different and various states of the market economy may symbolise different stages of the business cycle and current states of supply and demand, amongst the many other economic factors.

An associated challenge that arises from HMM calibration and implementation is the calculation of optimal model parameters. In this thesis, we contribute further to the literature by refining and extending the change of probability measure approach in Elliott et al. [49]. Recursive filters related to the underlying Markov chain are derived. The parameters are computed, in terms of the recursive filters, using the Expectation-Maximisation (EM) algorithm. Due to the recursive nature of estimation equations for each parameter, the estimates can be updated instantly as soon as new information becomes available in the market. This dynamic parameter estimation is a distinctive element in our implementation approach that makes all models proposed in the succeeding chapters self-tuning. A comprehensive overview of HMMs together with other approaches in parameter estimation and applications to other related fields can be found in Ephraim and Merhav [60].

The main objectives of this thesis are: (i) to develop several HMM-based models in capturing the dynamics of important market primitives; (ii) to calculate optimal parameter estimates of these models using HMM filtering techniques developed via

change of probability methods; (iii) to apply and assess these models within the three core aspects of financial modelling, which are pricing, asset allocation and risk measurement; and (iv) to investigate the forecasting performance of these models via error analysis and compare some of them to other known econometric models. To attain these objectives, this thesis is organised as follows. The first chapter gives a review of HMMs and an introduction to the change of measure technique which is applied subsequently to chapters 3, 5, 6 and 7. Chapter 3 puts forward an HMM-based model for commodity markets and takes on gold spot prices as a particular focus of study. The observation process is assumed to have drift and volatility, which are functions of a Markov chain in discrete time; furthermore, the volatility component has white noise perturbation. Adaptive and recursive filters are derived which in turn provide optimal model parameter estimates. A data set consisting of gold prices for the period 1973-2006 was utilised to illustrate the numerical computation involved in the implementation. We analyse the predictability of gold prices within the HMM framework using the Diebold-Killian metric. Our findings suggest that a two-state HMM is sufficient to describe the dynamics of the data and the gold price is predictable up to a certain extent in the short term but almost impossible to predict in the long term. Additionally, we benchmark the HMM forecasts with the forecasts of other known econometric models with respect to price predictability and forecasting errors.

A general filtering technique adopted from Elliott [48] is detailed in chapter 4. The recursive filters for the Markov chain and related processes are calculated and employed in the modelling set-up of chapters 5 and 6. In chapter 5, an HMM-based model for interest rate dynamics is proposed. In this setting, the observation process is governed by an Ornstein-Uhlenbeck (OU) process with Markov switching parameters. An on-line parameter estimation scheme using the EM algorithm is constructed. The proposed model is implemented on a data set of 30-day Canadian T-bills, which are used as a proxy for short term interest rates. We obtain one-step ahead forecasts that closely resemble the original time series data. That is, the regime-switching interest rate model developed yields very small prediction errors. A sensitivity analysis for the model with respect to changes in the parameter estimates is conducted with the aid of the Fisher information. Moreover, we

investigate the optimal number of states for the Markov chain using the Akaike information criterion.

Chapter 6 generalises the modelling set-up of chapter 5. This extension is essential to achieve a realistic model for electricity spot prices, which aside from mean-reversion also exhibit peculiar properties such as seasonality and frequently occurring jumps. The optimal estimates are derived recursively for the parameters of the OU-process and the jump size process. Once again, we have a self-calibrating model here, which is implemented on a deseasonalised series of daily spot electricity prices from the Nordic exchange Nord Pool. On the basis of one-step ahead forecasts, we found that the model is able to capture the empirical characteristics of Nord Pool spot prices. The pricing of expected spots on delivery is included to demonstrate an application of our model to pricing.

An exposition of a multivariate version of an HMM is presented in chapter 7. Here, a vector observation process is given with one underlying Markov chain. The optimal parameter estimation is described and as for the applications, we address two asset allocation problems. In the first problem, an investor is faced with the problem of competing investment strategies. The performance of these available investment strategies will be compared with that of the HMM-driven strategy which relies on the risk-adjusted forecasts. Under a two-dimensional HMM set-up, optimal allocation weights for a portfolio are further derived, where an investor can either invest in growth or value stocks. The second asset allocation problem deals with a stochastic programming type of enquiry, where an objective function needs to be optimised subject to the CVaR constraints. The HMM results and the scenario generation technique are interfaced in this chapter. It will be shown that the scenario generator gives stable solutions for a variety of portfolio optimisation problem formulations.

This thesis culminates with a pricing application of the HMM in chapter 8. Under a Markov-modulated Merton's structural model approach, the swap rate in a credit default swap (CDS) is calculated. The value of the firm has regime-switching parameters but in contrast to the previous chapters, the underlying Markov chain is

assumed observable and evolves in continuous time. The pricing method uses the Esscher transform to find an equivalent martingale measure. A system of coupled partial differential equations (PDEs) satisfied by the default probabilities under the risk-neutral and physical measures is derived. Under a two-state Markov chain, a numerical solution to this system of PDEs is obtained through the Crank-Nicolson scheme. A Monte-Carlo simulation is performed to compare the computational efficiency of the two numerical schemes. In the context of financial valuation, adding the possibility of regime shifts in a model can help overcome the underestimation of the default probabilities and CDS swap rates.

The thesis concludes in chapter 9 with the summary of the contributions of this research and the identification of possible specific directions, which are ramifications resulting from the analysis carried out in this work.

Chapter 2

Review of hidden Markov models

In this chapter the basic features of an HMM are explained. First, Markov chains are defined in a discrete-time setting. The second section gives a description of the general framework for HMMs and highlights their characteristics. A change of probability measure technique for HMMs is introduced in section 2.3. This technique is used in the succeeding chapters to derive adaptive filters and optimal parameter estimates.

2.1 Markov chains

A Markov process is a random process without memory. The future state of the process depends only on its current state, i.e., it is conditionally independent of the past. We follow the discussion in [123] and assume, that the Markov chain has a finite and countable set of states.

2.1.1 Markov chains in discrete time

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(\mathbf{x}_k)_{k\in\mathbb{N}}$ be a sequence of random variables with values in the state-space set $\mathcal{M} = \{\boldsymbol{m}_1, \boldsymbol{m}_2, ...\boldsymbol{m}_N\}$, where $\mathbf{x}$ is a function $\mathbf{x} : \Omega \rightarrow \mathcal{M}$ and $\mathbb{N}$ is the set of natural numbers.

Definition 2.1

The process $\mathbf{x}$ *is said to be a Markov chain, if it satisfies the Markov property*

$$P(\mathbf{x}_{k+1} = \boldsymbol{m}_{k+1} \mid \mathbf{x}_0 = \boldsymbol{m}_0, ..., \mathbf{x}_k = \boldsymbol{m}_k) = P(\mathbf{x}_{k+1} = \boldsymbol{m}_{k+1} \mid \mathbf{x}_k = \boldsymbol{m}_k) \quad (2.1)$$

$\forall k \geq 1$ *and* $\boldsymbol{m}_0, \boldsymbol{m}_1, ..., \boldsymbol{m}_k \in \mathcal{M}$.

The initial distribution of $\mathbf{x}$ is defined by $\chi = (\chi_m : m \in \mathcal{M})$, $\chi_{\boldsymbol{m}} = P(\mathbf{x} = \boldsymbol{m}) = P(\{\omega : \mathbf{x}(\omega) = \boldsymbol{m}\})$. Furthermore the Markov chain $(\mathbf{x}_k)_{k \in \mathbb{N}}$ is characterised by its transition probability matrix $\mathbf{\Pi}$. For a particular element π_{ji} of the transition probability matrix $\mathbf{\Pi}$,

$$\pi_{ji} = P(\mathbf{x}_{k+1} = \boldsymbol{j} \mid \mathbf{x}_k = \boldsymbol{i}), \quad \boldsymbol{i}, \boldsymbol{j} \in \mathcal{M} \quad (2.2)$$

where $\pi_{ji} \geq 0 \ \forall (\boldsymbol{j}, \boldsymbol{i}) \in \mathcal{M}^2$ and $\sum_{j \in \mathcal{M}} \pi_{ji} = 1 \, \forall i \in \mathcal{M}$.

These one-step transition probabilities $\pi'_{ji}s$ for the Markov chain denote the probability of switching from state i to state j. The Markov chain is time homogeneous, that means, that the transition probabilities $\pi_{ji} = P(\mathbf{x}_{k+1} = \boldsymbol{j} \mid \mathbf{x}_k = \boldsymbol{i})$ do not depend on time k.

The h-step ahead transition probabilities can be calculated by multiplying the matrix $\mathbf{\Pi}$ by itself h times. This matrix is denoted by $\mathbf{\Pi}^h$ and $\pi_{ji}^{(h)} = (\mathbf{\Pi}^h)_{ji}$ is the (j, i) entry in the h-step transition probability matrix $\mathbf{\Pi}^h$.

The states of a Markov chain can be represented by the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ of $\mathbb{R}^N$, where $\mathbf{e}_i = (0, ..., 0, 1, 0, ..., 0)^\top \in \mathbb{R}^k$, where $^\top$ denotes the transpose of the row vector. It is associated with the state space $\mathcal{M}$. When $\boldsymbol{m}_k = \boldsymbol{j}$, then the Markov chain $\mathbf{x}_k$ is represented by a unit vector with the element 1 in row j, and 0 elsewhere. The conditional expectation of $\mathbf{x}_{k+1}$ is then given by the jth column of the transition probability matrix,

$$\mathrm{E}(\mathbf{x}_{k+1} \mid \boldsymbol{m}_k = \boldsymbol{j}) = \begin{bmatrix} \pi_{j1} \\ \vdots \\ \pi_{jN} \end{bmatrix}$$

Therefore, we have

$$\mathrm{E}(\mathbf{x}_{k+1} \mid \mathbf{x}_k) = \mathrm{E}(\mathbf{x}_{k+1} \mid \mathbf{x}_k, \mathbf{x}_{k-1}, ...) = \mathbf{\Pi}\mathbf{x}_k \, . \quad (2.3)$$

In forecasting the states of a Markov chain, which are represented by the unit vectors, note that we can express the Markov chain in the form

$$\mathbf{x}_{k+1} = \mathbf{\Pi}\mathbf{x}_k + \mathbf{v}_{k+1} \tag{2.4}$$

where $\mathbf{v}_k$ is a martingale increment (see [81]). It is not possible to forecast $\mathbf{v}_k$ on the basis of previous states of the process and $\mathbf{v}_{k+1}$ can be seen to follow the difference

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \mathrm{E}(\mathbf{x}_{k+1} \mid \mathbf{x}_k, \mathbf{x}_{k-1}, ...)$$

The dynamics of $\mathbf{x}$ in formula (2.4) imply, that $\mathbf{x}_{k+h} = \mathbf{\Pi}^h\mathbf{x}_k + \mathbf{v}_{k+h} + \mathbf{\Pi}\mathbf{v}_{k+h-1} + \mathbf{\Pi}^2\mathbf{v}_{k+h-2} + ... + \mathbf{\Pi}^{h-1}\mathbf{v}_{k+1}$. Since $\mathbf{v}_k$ is a martingale increment, it follows, that the h-step ahead forecast of the Markov chain is given by

$$\mathrm{E}(\mathbf{x}_{k+h} \mid \mathbf{x}_k, \mathbf{x}_{k-1}, ...) = \mathbf{\Pi}^h\mathbf{x}_k \ . \tag{2.5}$$

2.2 Hidden Markov models

In an HMM a Markov chain is embedded in a stochastic process, which is a series of observations. The Markov chain itself is not observable, i.e., it is "hidden" in the observations. Our aim is to estimate the underlying Markov chain, that is, filter the sequence $\{\mathbf{x}_k\}$ out of the observations. The underlying Markov chain $\mathbf{x}_k$ is assumed to be homogeneous with finite state space in discrete time. Under the real world measure P, the Markov chain follows the dynamics $\mathbf{x}_{k+1} = \mathbf{\Pi}\mathbf{x}_k + \mathbf{v}_{k+1}$, where $\mathbf{\Pi}$ is the transition probability matrix and $\mathbf{v}_{k+1}$ is a martingale increment. The observation process is denoted by $\{y_k\}$ and can follow various types of dynamics.

HMMs were first introduced by Baum and Petrie in 1966 [13] and further developed in papers by Baum and his collaborators in the 60s and 70s (see [12], [15] and [11]). The Baum-Welch algorithm was introduced in 1970 by Baum et al. [14] for parameter estimation within HMMs, a short description of the algorithm is given in section 3.3.1. Further details of the development and applications of HMMs are

stated in Ephraim and Merhav [60] and the references therein.

The key elements involved in an HMM are described in a tutorial paper by Rabiner [129] which gives a thorough overview of HMMs. The first key element is the number of states N of the Markov chain. Since the Markov chain is hidden, the number of states is not observable. A reasonable choice has to be made on the basis of the observed process. A discussion of how to choose the optimal number of states is given at the end of chapter 5. The state space $\mathcal{M}$ is finite, more specifically $\mathcal{M} = \{\boldsymbol{m}_1, \boldsymbol{m}_2, ..., \boldsymbol{m}_N\}$, and as previously mentioned can be associated with the canonical basis of $\mathbb{R}^N$. In an ergodic HMM, all states are interconnected, therefore every state can be reached from any other state. The second element of an HMM is the number M of distinct observations. When the observation process is discrete, we select a set of distinct observations. The transition probability matrix $\boldsymbol{\Pi} = \{\pi_{ji}\}$ is the third element of an HMM and defines the state transition probability distribution, whilst the probability distribution of the observation process is the fourth key element. Finally, the HMM is characterised by the initial state distribution $\chi = \{\chi_{ji}\}$, where$\chi_j = P(\mathbf{x}_1 = \boldsymbol{m}_j)$, for all $1 \leq j \leq N$.

So far, the underlying Markov chain is defined on the real-world probability measure P. In the next section, a change of measure technique is introduced, which is applied and extended in later chapters for the derivation of filters for processes related to the underlying Markov chain.

2.3 Change of probability measure

This section gives a summary of a change of probability measure technique for our filtering problem. The change of measure technique is widely used in filtering applications and was introduced for stochastic filtering by Zakai [149]. Elliott et al. [49] utilises this change of measure technique, which is based on a discrete-time version of Girsanov's theorem to derive optimal filters. This technique enables us to make calculations under a mathematically "ideal" measure, which we shall term as a reference probability measure. This new "ideal" probability measure is equivalent to the real world measure, the measure under which we have the

observation process. The observations under the new measure are independent and identically random variables. The Markov chain follows the same dynamics under both the reference probability measure and the real world measure. Changing from the real measure to the ideal measure leads to easier ways of calculating filters as Fubini-type results can be employed instead of direct calculations that require hard semi-martingale methods.

2.3.1 Change of measure techniques

The theory of changing measures is based on the equivalence of two probability measures linked through the Radon-Nikodým derivative. Let $(\Omega, \mathcal{F})$ be a measurable space. We suppose P is a probability measure on $\mathcal{F}$. To construct an equivalent probability measure $\bar{P}$ on $(\Omega, \mathcal{F})$, we use the following theorem.

Theorem 2.2

If P and $\bar{P}$ are two probability measures with $P \ll \bar{P}$, then there exists a nonnegative function f, such that

$$P(A) = \int_A f \, d\bar{P} \;\; \forall A \in \mathcal{F}.$$

For two such functions f and g, $\bar{P}(f \neq g) = 0$, so the nonnegative function is unique.

Proof

See Billingsley [19].

□

With theorem 2.2 we can write

$$\int_A dP = \int_A \frac{dP}{d\bar{P}} d\bar{P} \;\; \forall A \in \mathcal{F}.$$

The measurable function $\frac{dP}{d\bar{P}}$ is the Radon-Nikodým derivative of P with respect to $\bar{P}$. The new probability measure $\bar{P}$ on $(\Omega, \mathcal{F})$ is defined via this Radon-Nikodým derivative (see Elliot et al. [49] for further discussion). We write

$$\left.\frac{d\bar{P}}{dP}\right|_{\mathcal{F}} := \Lambda,$$

so it follows, that

$$\bar{P}(A) = \int_A \Lambda \, dP \;\; \forall A \in \mathcal{F}.$$

To carry out the filter derivations for the processes of the Markov chain it is necessary to consider conditional expectations that relate the two equivalent measures, see Elliott et al [49]. The conditional Bayes' theorem, given in Appendix A, is fundamental in obtaining many major results of this thesis. For filtering applications which are discussed in the succeeding chapters we need a modified version of Bayes' theorem, namely the conditional Bayes' theorem for stochastic processes.

Write

$$\Lambda_t := \mathrm{E}[\Lambda \mid \mathcal{F}_t], \tag{2.6}$$

where Λ is the Radon-Nikodým derivative $\Lambda = \dfrac{d\bar{P}}{dP}$.

Theorem 2.3 (Conditional Bayes' theorem for stochastic processes)
Let $(\Omega, \mathcal{F}, P)$ be a probability space equipped with a complete filtration $\{\mathcal{F}_t\}$. Suppose $\bar{P}$ is another probability measure defined through the Radon-Nikodým derivative Λ. Let $\{\varpi_t\}$ be an $\mathcal{F}_t$-adapted process and Λ_t is the process in equation (2.6). Then

$$\bar{E}[\varpi_t \mid \mathcal{F}_t] = \frac{E[\Lambda_t \varpi_t \mid \mathcal{F}_s]}{E[\Lambda_t \mid \mathcal{F}_s]} \;\; 0 \le s \le t$$

provided $E[\Lambda_t \mid \mathcal{F}_s] > 0$.

Proof

The proof of this theorem is similar to the proof of theorem A.2 in Appendix A.

□

Collorary 2.4

The process Λ_t is a martingale.

Proof

$$E[\Lambda_{t+1} \mid \mathcal{F}_t] = E\Big[E[\Lambda \mid \mathcal{F}_{t+1}] \mid \mathcal{F}_t\Big] \overset{\text{tower property}}{=} E[\Lambda \mid \mathcal{F}_t] = \Lambda_t$$

□

2.3.2 Change of measure for discrete time processes

Under the discrete time framework, let $\{X_k\}$, $k \in \mathbb{N}$ be a sequence of random variables with positive probability density functions (pdf) ϕ_k on $(\Omega, \mathcal{F}, P)$. Corresponding to this sequence, we have the filtration $\{\mathcal{F}_k\}$ generated by $\sigma\{X_1, ..., X_k\}$. We wish to define a new probability measure $\bar{P}$ on $(\Omega, \bigcup_{k\geq 0} \mathcal{F}_k)$, such that $\{X_k\}$ is independent and identically distributed (IID) with positive pdf α. To attain this objective, define

$$\lambda_0 := 1,\ \lambda_l := \frac{\alpha(X_l)}{\phi_l(X_l)} \text{ for } l \geq 1 \text{ and } \Lambda_k := \prod_{l=0}^{k} \lambda_l\,.$$

Write

$$\frac{d\bar{P}}{dP}(\omega)\bigg|_{\mathcal{F}_k} := \Lambda_k(\omega)\,.$$

Lemma 2.5

The random variables Λ_k, $k \geq 0$ are P-martingales under $\mathcal{F}_k$ and $E[\Lambda_k] = 1$. Furthermore, under $\bar{P}$, $\{X_k\}$ is a sequence of IID random variables with pdf α.

Proof

First, we show that $E[\Lambda_k \mid \mathcal{F}_{k-1}] = \Lambda_{k-1}$.

$$\begin{aligned} E[\Lambda_k \mid \mathcal{F}_{k-1}] &= E[\Lambda_{k-1}\lambda_k \mid \mathcal{F}_{k-1}] \\ &= \Lambda_{k-1}E[\lambda_k \mid \mathcal{F}_{k-1}] \quad \textit{(since } \Lambda_{k-1} \textit{ is } \mathcal{F}_{k-1} \textit{ measurable)} \\ &= \Lambda_{k-1}E\left[\frac{\alpha(x_k)}{\phi_k(x_k)}\middle|\mathcal{F}_{k-1}\right]. \end{aligned}$$

For all $F \in \mathcal{F}_{k-1}$ we have

$$\begin{aligned} E\left[\frac{\alpha(x_k)}{\phi_k(x_k)} \mid \mathcal{F}_{k-1}\right] &= \int_{-\infty}^{\infty} \frac{\alpha(x_k)}{\phi_k(x_k)}\phi_k(x_k)\,dx_k \\ &= \int_{-\infty}^{\infty} \alpha(x_k)\,dx_k = 1\,. \end{aligned}$$

It is easy to see that $E[\Lambda_k] = 1$ *by noting that*

$$\begin{aligned} E[\Lambda_k] &= E[\Lambda_k \mid \mathcal{F}_0] = \Lambda_0 \quad \textit{(since } \Lambda_k \textit{ is a martingale)} \\ &= \lambda_0 = 1\,. \end{aligned}$$

For any $\mathcal{F}$*-measurable function* g

$$\begin{aligned} \bar{E}[g(x_k) \mid \mathcal{F}_{k-1}] &= \frac{E[\Lambda_k g(x_k) \mid \mathcal{F}_{k-1}]}{E[\Lambda_k \mid \mathcal{F}_{k-1}]} \quad \textit{(by Bayes' theorem 2.3)} \\ &= \frac{\Lambda_{k-1} E[\lambda_k g(x_k) \mid \mathcal{F}_{k-1}]}{\Lambda_{k-1}} \\ &= E[\lambda_k g(x_k) \mid \mathcal{F}_{k-1}] = E\left[\frac{\alpha(x_k)}{\phi(x_k)} g(x_k) \mid \mathcal{F}_{k-1}\right] \\ &= \int_{\mathbb{R}} \frac{\alpha(x_k)}{\phi(x_k)} \phi(x_k)\, dx_k = \int_{\mathbb{R}} g(x_k)\alpha(x_k)\, dx_k \end{aligned}$$

So, under $\bar{P}$, $\{X_k\}$ *is a sequence of IID random variables with pdf* α.

□

2.3.3 Change of measure for linear systems

The preceding theory is now applied to a change of measure for linear systems. This general change of measure technique for normally distributed random variables will be applied in the next chapter. In chapter 5 and 6, this technique is extended to deal with more complex underlying processes.

We begin the change of measure with processes defined on an ideal probability space $(\Omega, \mathcal{F}, \bar{P})$. Let $\{\nu_k\}$, $k \in \mathbb{N}$ be a sequence of IID N(0,1) random variables under $\bar{P}$. Define

$$\nu_{k+1} := b^{-1}(X_{k+1} - aX_k)\,.$$

so that the state of the system is described by the linear dynamics

$$X_{k+1} = aX_k + b\nu_{k+1}\,. \tag{2.7}$$

Here, $\mathrm{E}[\nu_{k+1} \mid \mathcal{F}_t] = 0$. It is immediate to see that $\{\nu_k\}$, $k \in \mathbb{N}$ is a sequence of martingale increments. We wish to define a measure P, such that the dynamics in equation (2.7) will hold.

Let $\bar{P}$ be a probability measure on $(\Omega, \mathcal{F})$. Under $\bar{P}$, $\{X_k\}, k \in \mathbb{N}$, is IID $N(0,1)$ with density ϕ. For $l = 0, 1, 2...$ and $b \neq 0$ define

$$\begin{aligned} \bar{\lambda}_l &= \frac{\phi\left(b^{-1}(X_l - aX_{l-1})\right)}{b\phi(X_l)} \\ \bar{\Lambda}_k &= \prod_{l=0}^{k} \bar{\lambda}_l \,. \end{aligned}$$

Lemma 2.6

The process $\bar{\Lambda}_k, k \in \mathbb{N}$ is a $\bar{P}$-martingale with respect to the filtration $\{\mathcal{F}_k\}$.

Proof

$$\begin{aligned} \bar{E}[\bar{\Lambda}_{k+1} \mid \mathcal{F}_k] &= \bar{E}[\bar{\Lambda}_k \bar{\lambda}_{k+1} \mid \mathcal{F}_k] = \bar{\Lambda}_k \bar{E}[\bar{\lambda}_{k+1} \mid \mathcal{F}_k] \\ &= \bar{\Lambda}_k \int_{\mathbb{R}} \frac{\phi(b^{-1}(X_{k+1} - aX_k))}{b\phi(x_{k+1})} \phi(x_{k+1})\, dx_{k+1} \\ &= \bar{\Lambda}_k \int_{\mathbb{R}} \phi(\nu_{k+1}) d\nu_{k+1} = \bar{\Lambda}_k \,. \end{aligned}$$

□

Now we define a new probability measure P on $(\Omega, \mathcal{F})$ by setting the restriction of the Radon-Nikodým derivative to $\mathcal{F}_k$ equal to $\bar{\Lambda}_k$. That is,

$$\left.\frac{dP}{d\bar{P}}\right|_{\mathcal{F}_k} = \bar{\Lambda}_k \,.$$

Lemma 2.7

On $(\Omega, \mathcal{F})$ and under P, $\{\nu_k\}$ $k \in \mathbb{N}$, is a sequence of IID $N(0,1)$ random variables with

$$\nu_{k+1} := b^{-1}(X_{k+1} - aX_k) \,. \tag{2.8}$$

Proof

Let g be an $\mathcal{F}$-measurable function on $(\Omega, \mathcal{F})$. We have to show that

$$E[g(\nu_{k+1}) \mid \mathcal{F}_k] = \int_{\mathbb{R}} g(x)\phi(x)\, ds \,.$$

$$
\begin{aligned}
E[g(\nu_{k+1}) \mid \mathcal{F}_k] &= \frac{\bar{E}\left[\bar{\Lambda}_{k+1} g(\nu_{k+1}) \mid \mathcal{F}_k\right]}{\bar{E}\left[\bar{\Lambda}_{k+1} \mid \mathcal{F}\right]} \\
&= \frac{\bar{\Lambda}_k \bar{E}\left[\bar{\lambda}_{k+1} g(\nu_{k+1}) \mid \mathcal{F}\right]}{\bar{\Lambda}_k} \quad \textit{since } (\bar{\Lambda})_k \textit{ is a martingale} \\
&= \bar{E}\left[\bar{\lambda}_{k+1} g(\nu_{k+1}) \mid \mathcal{F}_k\right] \\
&= \bar{E}[\frac{\phi(\nu_{k+1})}{b\phi(x_{k+1})} g(\nu_{k+1}) \mid \mathcal{F}_k] \\
&= \int_{-\infty}^{\infty} \frac{\phi(\nu_{k+1})}{b\,\phi(x_{k+1})} g(\nu_{k+1}) \phi(x_{k+1})\, dx_{k+1} \\
&= \int_{-\infty}^{\infty} \frac{1}{b} \phi(\nu_{k+1}) g(\nu_{k+1})\, dx_{k+1} \\
&= \int_{-\infty}^{\infty} \phi(\nu_{k+1}) g(\nu_{k+1})\, d\nu_{k+1} \ .
\end{aligned}
$$

□

Therefore a new probability measure P on $(\Omega, \mathcal{F})$ is found, which is equivalent to the measure of the 'ideal' world $\bar{P}$. This enables a change of measure back to the 'real' world, in which the observation process $\{y_k\}$ is observed and optimal parameter estimates are needed.

Chapter 3

Modelling commodity prices A hidden Markov model for gold prices

In this chapter an HMM methodology to model gold spot prices is developed. Adaptive filters are derived which are used for recursive parameter estimates of the model. We start this chapter by formulating an HMM for commodities in section 3.1. In section 3.2 the adaptive and recursive filters are derived which are used for the estimation of the model parameters in section 3.3. Although the filtering algorithm was described in Elliott et al. [49], we refine and develop in this chapter a more compact form of the closed-form solutions of the recursive filters given in theorem 3.2. In our formulation, the recursions all involve neater expressions containing only matrices and vectors, whereas in Elliott et al. [49] the recursions are in the form of summations.

The model is implemented on a data set in section 3.4 and a forecastability analysis is given in section 3.5. The forecastability performance of the model is compared to that of well-known time series models, namely autoregressive conditional heteroscedasticity (ARCH) and generalised autoregressive conditional heteroscedasticity (GARCH) models in section 3.6.

3.1 Hidden Markov model for commodities

To develop a hidden Markov model for spot gold prices we first take a look at one of the popular models for spot prices of commodities, the Gibson-Schwartz model [73]. This model was originally developed for pricing contingent claims on oil. It is a two-factor model, where the spot price of oil and the spot instantaneous convenience yield follow a joint stochastic process. In the case of oil, the convenience yield has a considerable impact on the dynamics of the spot price. Carmona and Ludkovski [30] define the convenience yield as the difference between the benefit of direct access to a commodity and the cost of carry. The convenience yield is unobserved in the market and thus it is modelled indirectly to correct the drift of the spot price process.

Whilst the Gibson-Schwartz model shows a good performance in the empirical implementation (Carmona & Ludkovski [30], Schwartz [134]) for the oil or the copper market, gold data has specific characteristics that must be considered to model the gold spot prices accurately. As discussed in Schwartz [134], the mean reversion in the convenience yield as well as the mean reversion of the commodity price are insignificant for gold. Therefore it is possible to assume a convenience yield of zero in modelling the behaviour of gold prices.

On the premise that a convenience yield is not really significant for gold, a one-factor model is deemed sufficient as a basis for the hidden Markov model. Let us consider the one-factor model of Brennan and Schwartz [26], which was proposed for the evaluation of natural resources. A commodity price model based on a geometric Brownian motion was also suggested by Miltersen [117]. The spot price of a commodity is modelled by a geometric Brownian motion (GBM) according to the specification

$$dS_t = fS(t)\,dt + \sigma S(t)\,dW_t$$

where W is a Wiener process, σ is the instantaneous standard deviation of the spot price and f is the drift of the stochastic process. So, by Itô's lemma, S has the solution

$$S_t = S_0 \exp\left[(f - \frac{\sigma^2}{2})t + \sigma W_t\right].$$

The GBM dynamics are our starting point in building the hidden Markov model for gold prices. Suppose $(\Omega, \mathcal{F}, P)$ denotes a probability space under which $\mathbf{x}_k$ is a homogeneous Markov chain with finite state space in discrete time $(k = 0, 1, 2...)$. The state space of $\mathbf{x}_k$ is associated with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\} \in \mathbb{R}^N$ with $\mathbf{e}_i = (0, ..., 0, 1, 0, ..., 0)^\top \in \mathbb{R}^n$ (see Elliot et al. [49]). The initial distribution of $\mathbf{x}_0$ is known and $\mathbf{\Pi} = (\pi_{ji})$ is the transition probability matrix with $\pi_{ji} = P(\mathbf{x}_{k+1} = e_j | \mathbf{x}_k = e_i)$. Let $\mathcal{F}_k^{\mathbf{x}_0} = \sigma\{\mathbf{x}_0, ..., \mathbf{x}_k\}$ be the σ-field generated by $\mathbf{x}_0, ..., \mathbf{x}_k$ and let $\mathcal{F}_k^{\mathbf{x}}$ be the complete filtration generated by $\mathcal{F}_k^{\mathbf{x}_0}$.

The Markov chain $\mathbf{x}_k$ in this model for gold prices is not directly observable, but the process is "hidden" in the logarithmic returns y_{k+1} of the spot price S_k for gold. So we observe a function $g(\mathbf{x}_k, w_{k+1})$

$$y_{k+1} = \ln \frac{S_{k+1}}{S_k} = g(\mathbf{x}_k, w_{k+1}) \tag{3.1}$$

where $\mathbf{x}_k$ has finite state space and w_k's constitute a sequence of IID random variables independent of $\mathbf{x}$. The real-valued process y is assumed to satisfy the equation

$$y_{k+1} = f(\mathbf{x}_k) + \sigma(\mathbf{x}_k) w_{k+1} = \langle \boldsymbol{f}, \mathbf{x}_k \rangle + \langle \boldsymbol{\sigma}, \mathbf{x}_k \rangle \, w_{k+1} \,. \tag{3.2}$$

Note that $\mathbf{f} = (f_1, f_2, ..., f_N)^\top$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_N)^\top$ are vectors, $f(\mathbf{x}_k) = \langle \mathbf{f}, \mathbf{x}_k \rangle$ and $\sigma(X_k) = \langle \boldsymbol{\sigma}, \mathbf{x}_k \rangle$, where $\langle \mathbf{b}, \mathbf{c} \rangle$ denotes the Euclidean scalar product in $\mathbb{R}^N$ of the vectors $\mathbf{b}$ and $\mathbf{c}$. We assume $\sigma_i \neq 0$. Let $\mathcal{F}_k^y$ be the filtration generated by the $\sigma(y_1, y_2, ..., y_k)$ and $\mathcal{F}_k = \mathcal{F}_k^{\mathbf{x}} \vee \mathcal{F}_k^y$ is the global filtration.

Thus the hidden Markov model for the gold price process S_k is based on the model by Brennan & Schwartz [26] and is extended in this thesis through the inclusion of an N-state Markov chain. The states of the Markov chain can represent different states of the economy. If we assume $N = 3$, the Markov chain can refer perhaps to a 'booming', 'medium' or 'recessive' economy. The observation at time $k + 1$ depends on the state of $\mathbf{x}$ at time k. It is a one-step delay model, which indicates, that the price process does not react immediately to a change of the economic state.

The derivation of the filters for the Markov chain shall be done under a reference probability measure $\bar{P}$, which is equivalent to the real measure P. Under the real world probability measure P, $\mathbf{x}$ has the dynamics

$$\mathbf{x}_{k+1} = \Pi \mathbf{x}_k + \mathbf{v}_{k+1} \tag{3.3}$$

where $\mathbf{v}_{k+1}$ is a martingale increment and $\Pi = (\pi_{ji})$ is the transition probability matrix. Under $\bar{P}$ X is still a Markov chain with $\mathbf{x}_{k+1} = \Pi \mathbf{x}_k + \mathbf{v}_{k+1}$ where $\mathbf{v}_{k+1}$ is the martingale increment with $\bar{E}[\mathbf{v}_{k+1} \mid \mathcal{F}_k^{\mathbf{x}}] = \mathbf{0}$.

Following the change of measure technique which was outlined in chapter 2 the adaptive filters for the Markov chain are derived under an "idealised" measure $\bar{P}$. The real world measure P is constructed from $\bar{P}$ through the Radon-Nikodým derivative of P with respect to $\bar{P}$, $\frac{dP}{d\bar{P}}\Big|_{\mathcal{F}_k} = \Lambda_k$. To construct Λ_k we define the process λ_l

$$\lambda_l := \frac{\phi\Big[\sigma(\mathbf{x}_{l-1})^{-1}\big(y_l - f(\mathbf{x}_{l-1})\big)\Big]}{\sigma(\mathbf{x}_{l-1})\phi(y_l)} \tag{3.4}$$

where $\phi(z)$ is the probability density function of a standard normal random variable Z and set

$$\Lambda_k := \prod_{l=1}^{k} \lambda_l, \quad k \geq 1, \quad \Lambda_0 = 1\,.$$

From lemma 2.7 we know that under P the sequence of variables $z_1, z_2, \ldots,$ is a sequence of IID standard normals where $z_{k+1} = \sigma(\mathbf{x}_k)^{-1}\left(y_{k+1} - f(\mathbf{x}_k)\right)$.

Our aim is to estimate the Markov chain $\mathbf{x}$, given the observations under P, the real world probability. As argued before it is easier to perform the calculations under $\bar{P}$. Let us characterise the conditional distribution of $\mathbf{x}_k$ given $\mathcal{F}_k$ under P. Write

$$\widehat{p}_k^{\,i} := P(\mathbf{x}_k = e_i | \mathcal{F}_k^y) = \mathrm{E}\left[\langle \mathbf{x}_k, e_i \rangle \,|\mathcal{F}_k^y\right]$$

with $\widehat{\mathbf{p}}_k = (\widehat{p}_k^1, \widehat{p}_k^2, \ldots, \widehat{p}_k^N)^\top \in I\!R^N$. With Bayes' theorem 2.3 we see that

$$\widehat{\mathbf{p}}_k = \mathrm{E}[\mathbf{x}_k | \mathcal{F}_k^y] = \frac{\bar{\mathrm{E}}[\Lambda_k \mathbf{x}_k | \mathcal{F}_k^y]}{\bar{\mathrm{E}}[\Lambda_k | \mathcal{F}_k^y]}.$$

Now we define $\Xi_k := \bar{\mathrm{E}}[\Lambda_k \mathbf{x}_k | \mathcal{F}_k^y]$. Recall that $\sum_{i=1}^N \langle \mathbf{x}_k, e_i \rangle = 1$. Consequently,

$$\begin{aligned} \sum_{i=1}^N \langle \Xi_k, e_i \rangle &= \sum_{i=1}^N \bar{\mathrm{E}}\left[\langle \Lambda_k \mathbf{x}_k, e_i \rangle\right] = \bar{\mathrm{E}}\left[\Lambda_k \sum_{i=1}^N \langle \mathbf{x}_k, e_i \rangle \middle| \mathcal{F}_k^y\right] \\ &= \bar{\mathrm{E}}\left[\Lambda_k | \mathcal{F}_k^y\right]. \end{aligned} \tag{3.5}$$

With (3.5) and the definition of Ξ_k we have

$$\widehat{\mathbf{p}}_k = \frac{\Xi_k}{\sum_{i=1}^N \langle \Xi_k, e_i \rangle}.$$

3.2 Adaptive and recursive filters

The aim of this section is to derive adaptive and recursive filters for the underlying Markov chain. The adaptive filters enable coefficients to adjust to current market situations. This adjustment is achieved with the help of a recursive algorithm within the filter. Consequently, a "self-tuning" model is created, which adapts itself to changes in the time series data. In a recursive filter previous output values from the filter are used as inputs for the calculations.

First recursive filters for the conditional expectation $\Xi_k = \bar{\mathrm{E}}[\Lambda_k \mathbf{x}_k | \mathcal{F}_k^y]$ are computed. Let $\boldsymbol{D}_{k+1}$ be a diagonal matrix whose entries d_{ij} are defined by

$$d_{ij} = \begin{cases} \frac{\phi\left(\frac{y_{k+1} - f_i}{\sigma_i}\right)}{\sigma_i \phi(y_{k+1})} & \text{for } i = j \\ 0 & \text{otherwise} \end{cases}.$$

The entries of the diagonal matrix $\boldsymbol{D}_{k+1}$ for the case $i = j$ are the componentwise elements from the process λ_l in equation (3.4) defined for the Radon-Nikodým derivative.

Theorem 3.1

The recursive filter for Ξ_k is

$$\Xi_{k+1} = \boldsymbol{\Pi} \boldsymbol{D}_{k+1} \Xi_k. \tag{3.6}$$

Proof

From the definition of Ξ_k, we have $\Xi_{k+1} = \bar{E}\left[\Lambda_{k+1} \mathbf{x}_{k+1} | \mathcal{F}_{k+1}^y\right]$. With the dynamics

of $\mathbf{x}_{k+1}$ *in equation* (3.3), *we obtain*

$$\begin{aligned}
&\bar{E}\left[\Lambda_{k+1}\mathbf{x}_{k+1}|\mathcal{F}_{k+1}^{y}\right] \\
&= \bar{E}\left[\Lambda_k\lambda_{k+1}(\mathbf{\Pi}\mathbf{x}_k + \mathbf{v}_{k+1})|\mathcal{F}_{k+1}^{y}\right] \\
&= \sum_{i=1}^{N} \bar{E}\left[\Lambda_k \left\langle \mathbf{x}_k, \mathbf{e}_i\right\rangle |\mathcal{F}_k^{y}\right] \times \lambda_{k+1}\mathbf{\Pi}\mathbf{e}_i \qquad \textit{since } \bar{E}[\mathbf{v}_{k+1}] = \mathbf{0} \in I\!R^N \\
&= \sum_{i=1}^{N} \bar{E}\left[\Lambda_k \left\langle \mathbf{x}_k, \mathbf{e}_i\right\rangle |\mathcal{F}_k^{y}\right] \times \frac{\phi(\sigma_i^{-1}(y_{k+1} - f_i))}{\sigma_i\phi(y_{k+1})}\mathbf{\Pi}\mathbf{e}_i \\
&\quad \textit{using the definition of } \lambda_l \\
&= \sum_{i=1}^{N} \left\langle \mathbf{e}_i, \Xi_k\right\rangle \frac{\phi(\sigma_i^{-1}(y_{k+1} - f_i))}{\sigma_i\phi(y_{k+1})}\mathbf{\Pi}\mathbf{e}_i = \mathbf{\Pi}\boldsymbol{D}_{k+1}\Xi_k.
\end{aligned}$$

□

To obtain optimal recursive estimators for the parameters of the model we first have to analyse the Markov chain $\mathbf{x}_k$ with the dynamics $\mathbf{x}_{k+1} = \mathbf{\Pi}\mathbf{x}_k + \mathbf{v}_{k+1}$. We consider three processes following the exposition in Elliott et al. [49].

Consider the number of jumps of a Markov chain from state r to state s in time k defined by

$$J_k^{(sr)} := \sum_{l=1}^{k} \left\langle \mathbf{x}_{l-1}, \mathbf{e}_r\right\rangle \left\langle \mathbf{x}_l, \mathbf{e}_s\right\rangle .$$

Secondly, consider the occupation time, that is the length of time $\mathbf{x}$ has spent in state r up to time k. This is given by

$$O_k^{(r)} := \sum_{l=1}^{k} \left\langle \mathbf{x}_{l-1}, \mathbf{e}_r\right\rangle \ .$$

An auxiliary process is also needed to estimate the vectors $\mathbf{f} = (f_i)$ and $\boldsymbol{\sigma} = (\sigma_i)$ and this has the form

$$T_k^{(r)}(g) := \sum_{l=1}^{k} \left\langle \mathbf{x}_{l-1}, \mathbf{e}_r\right\rangle g(y_l)$$

where g is a function, which is either $g(y) = y$ or $g(y) = y^2$.

For any $\mathcal{F}^y$-adapted process H_k we shall write $\widehat{H}_k := \mathrm{E}[H_k|\mathcal{F}_k^y]$. The conditional expectation of H_k given $\mathcal{F}_k^y$ is denoted by $\eta_k(H_k) := \bar{\mathrm{E}}[\Lambda_k H_k|\mathcal{F}_k^y]$. We wish to develop recursive relations for conditional expectations of the processes defined above.

Applying the Bayes' theorem we consider

$$\widehat{J}_k^{(sr)} = \mathrm{E}[J_k^{(sr)}|\mathcal{F}_k^y] = \frac{\bar{\mathrm{E}}[\Lambda_k J_k^{(sr)}|\mathcal{F}_k^y]}{\bar{\mathrm{E}}[\Lambda_k|\mathcal{F}_k^y]}.$$

Although we cannot find a recursive expression for $\bar{\mathrm{E}}[\Lambda_k J_k^{(sr)}|\mathcal{F}_k^y]$ we find one for the vector process $\bar{\mathrm{E}}[\Lambda_k J_k^{(sr)}\mathbf{x}_k|\mathcal{F}_k^y] = \eta_k(J_k^{(sr)}\mathbf{x}_k)$.

The recursive relations for $\eta_k(J_k^{(sr)}\mathbf{x}_k)$, $\eta_k(O_k^{(r)}\mathbf{x}_k)$ and $\eta_k(T_k^{(r)}\mathbf{x}_k)$ are given in the following theorem.

Theorem 3.2

If $\boldsymbol{D}$ *is the diagonal matrix defined above then*

$$\begin{aligned}
\eta_l(J_l^{(sr)}\mathbf{x}_l) &= \mathbf{\Pi}\boldsymbol{D}_l(y_l)\eta_{l-1}(J_{l-1}^{(sr)}\mathbf{x}_{l-1}) \\
&\quad + \langle \Xi_{l-1}, \mathbf{e}_r \rangle \frac{\phi\left(\sigma_r^{-1}(y_l - f_r)\right)}{\sigma_r \phi(y_l)} \pi_{sr}\mathbf{e}_s \qquad (3.7) \\
\eta_l(O_l^{(r)}\mathbf{x}_l) &= \mathbf{\Pi}\boldsymbol{D}_l(y_l)\eta_{l-1}(O_{l-1}^{(r)}\mathbf{x}_{l-1}) \\
&\quad + \langle \Xi_{l-1}, \mathbf{e}_r \rangle \frac{\phi\left(\sigma_r^{-1}(y_l - f_r)\right)}{\sigma_r \phi(y_l)} \mathbf{\Pi}\mathbf{e}_r \qquad (3.8)
\end{aligned}$$

and

$$\begin{aligned}
\eta_l(T_l^{(r)}(g)\mathbf{x}_l) &= \mathbf{\Pi}\boldsymbol{D}_l(y_l)\eta_{l-1}(T_{l-1}^{(r)}(g)\mathbf{x}_{l-1}) \\
&\quad + \langle \Xi_{l-1}, \mathbf{e}_r \rangle \frac{\phi(\sigma_r^{-1}(y_l - f_r))}{\sigma_r \phi(y_l)} g(y_l)\mathbf{\Pi}\mathbf{e}_r \ . \qquad (3.9)
\end{aligned}$$

Proof

See Appendix B.1.

□

3.3 Model parameter estimates

The recursive relations for the processes of the Markov chain derived in theorem 3.2 are needed to derive optimal recursive estimators for the parameters of the model, which are the

1) transition matrix $\mathbf{\Pi} = (\pi_{ji})$,

2) drift parameter $\mathbf{f} = (f_i)$ and

3) variance vector $\boldsymbol{\sigma} = (\sigma_i)$.

The optimal parameter estimates are obtained through the technique of maximum likelihood estimation (MLE). The method used to calculate the MLE's is the so-called Expectation-Maximization (EM) algorithm, which is described below.

3.3.1 The EM algorithm

In this subsection we describe the EM algorithm. This technique is applied to derive optimal parameter estimates for the model parameter set $\widehat{\theta} = \{\widehat{\pi}_{ji}, \widehat{f}_i, \widehat{\sigma}_i, 1 \leq i, j \leq n\}$. The EM algorithm is an iterative procedure to find the MLE in incomplete data problems, where the computation of MLE's would be difficult because of the missing values or where optimisation of the likelihood function is analytically intractable (see McLachlan [114]). The EM algorithm was first developed by Dempster, Laird and Rubin [42] and has been widely used in engineering, computing and economics.

Let θ be a set of parameters in the parameter space Θ. Let $\{P^\theta, \theta \in \Theta\}$ be a family of probability measures on a measurable space $(\Omega, \mathcal{F})$, which is absolutely continuous with respect to a fixed probability measure P^0 and let $\mathcal{Y} \subset \mathcal{F}$.

We aim to calculate an optimal estimate of θ. The likelihood function for calculating θ on the basis of information contained in $\mathcal{Y}$ is given by

$$F(\theta) = E^0\left[\frac{dP^\theta}{dP^0}\middle| \mathcal{Y}\right]$$

and the maximum likelihood estimator (MLE) of θ is defined by

$$\widehat{\theta} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \; F(\theta)\,.$$

However, the MLE is not straightforward to compute. The EM algorithm approaches the problem indirectly with an iterative approximation method (see Elliott and Krishnamurthy [55] for a review).

We set $m = 0$ and choose $\widehat{\theta}_0$. For each iteration the EM algorithm consists of two steps: the expectation (E) step and the maximisation (M) step.

- In the Expectation step set $\theta^* = \widehat{\theta}_m$ and determine the function $Q(\theta, \theta^*)$

$$Q(\theta, \theta^*) = E^{\theta^*}\left[\log \frac{dP^{\theta}}{dP^{\theta^*}} \middle| \mathcal{Y}\right].$$

- In the Maximisation step find any value of $\theta \in \Theta$ that maximises $Q(\theta, \theta^*)$, that is

$$\widehat{\theta}_{m+1} \in \underset{\theta \in \Theta}{\operatorname{argmax}}\ Q(\theta, \theta^*) \ .$$

Lastly, replace m by $m + 1$ and repeat the E- and M-steps until some stopping criterion is satisfied.

As shown in Wu [147], the sequence $\{\widehat{\theta}_m\}$ yields non-decreasing values of the likelihood function that converge to a local maximum of the likelihood function. A well-known algorithm for parameter estimation in HMMs is the Baum-Welch algorithm [14]. This algorithm is a particular instance of the EM algorithm generalised for HMMs. It is a forward-backward algorithm which calculates the forward and backward probabilities for each state of the HMM and uses these probabilities to compute the MLEs of the parameters (see Cappé et al. [29] for further details on this algorithm). Now in our model the parameters can be optimised by applying the EM algorithm to $\log \frac{dP^{\theta}}{dP^{\widehat{\theta}}}$ with the previously defined set $\widehat{\theta}$. The aim is to find $\widehat{\theta}$ that maximises the analogues of the Q function.

3.3.2 Optimal parameter estimates

We perform a change of measure as described in subsection 2.3.1. Under P^{θ}, $\mathbf{x}$ is a Markov chain with transition matrix $\mathbf{\Pi} = (\pi_{ji})$. We shall introduce a new probability measure $P^{\widehat{\theta}}$ and under $P^{\widehat{\theta}}$, $\mathbf{x}$ is a Markov chain with transition matrix $\widehat{\mathbf{\Pi}} = (\widehat{\pi}_{ji})$. In other words, $P^{\widehat{\theta}}(\mathbf{x}_{k+1} = \mathbf{e}_j | \mathbf{x}_k = \mathbf{e}_i) = \widehat{\pi}_{ji}$. Thus, $\widehat{\pi}_{ji} \geq 0$ and $\sum_{j=1}^{n} \widehat{\pi}_{ji} = 1$.

In this situation,

$$\left.\frac{dP^{\widehat{\theta}}}{dP^{\theta}}\right|_{\mathcal{F}_k} = \Lambda_k,\ \Lambda_0 = 1 \text{ and } \Lambda_k = \prod_{l=1}^{k}\left(\prod_{s,r=1}^{n}\left(\frac{\widehat{\pi}_{sr}}{\pi_{sr}}\right)^{\langle \mathbf{x}_l, \mathbf{e}_s\rangle\langle \mathbf{x}_{l-1}, \mathbf{e}_r\rangle}\right). \tag{3.10}$$

See lemma 7.1, p. 37 of [49] for the justification of equation (3.10). When $\pi_{ji} = 0$, take $\widehat{\pi}_{ji} = 0$ and $\frac{\widehat{\pi}_{ji}}{\pi_{ji}} = 1$. The optimal estimates for the model parameters are given by the following theorem.

Theorem 3.3

If a sequence of observations $y_1, y_2, \ldots, y_k$ is available at time k, and the set of parameters $\{\widehat{\pi}_{ji}, \widehat{f}_i, \widehat{\sigma}_i\}$ determines the model then the filter EM estimates for these parameters are given by

$$\widehat{\pi}_{ji} = \frac{\widehat{J}_k^{(ji)}}{\widehat{O}_k^{(i)}} = \frac{\eta_k(J_k^{(ji)})}{\eta_k(O_k^{(i)})} \tag{3.11}$$

$$\widehat{f}_i = \frac{\widehat{T}_k^{(i)}}{\widehat{O}_k^{(i)}} = \frac{\eta_k(T_k^i(y))}{\eta_k(O_k^{(i)})} \tag{3.12}$$

and

$$\widehat{\sigma}_i = \sqrt{\frac{\widehat{T}_k^{(i)}(y^2) - 2f_i\widehat{T}_k^{(i)}(y) + f_i^2\widehat{O}_k^{(i)}}{\widehat{O}_k^{(i)}}}$$

$$= \sqrt{\frac{\eta_k(T_k^{(i)}y^2) - 2f_i\eta_k(T_k^{(i)}(y)) + f_i^2\eta_k(O_k^{(i)})}{\eta_k(O_k^{(i)})}}. \tag{3.13}$$

Proof

See Appendix B.2.

□

The above results not only provide estimates of the Markov chain but also update the parameters of the model. Recall from formulae (3.1) and (3.2) that the hidden Markov model for the logarithmic returns of daily gold prices S_k is of the form

$$y_{k+1} = \ln\frac{S_{k+1}}{S_k} = f(\mathbf{x}_k) + \sigma(\mathbf{x}_k)w_{k+1}.$$

The parameters f_i and σ_i are governed by a Markov chain $\mathbf{x}$ with N states, where $\mathbf{x}$ is not directly observable. Through the optimal parameter estimates derived in

theorem 3.3, these parameters can be updated whenever new information of the Markov chain is available in the market.

3.4 Implementation of filters to a data set

The recursive filters for the parameter estimation of the previous section are implemented on a time series of daily spot prices for gold. The recursive filters are tested on the log returns series of daily gold prices recorded from 1973 to 2006. The data set is based on London Afternoon Gold Price Fixes consisting of 8420 data values (available online at www.lbma.org.uk) which were retrieved from the London Bullion Market Association.

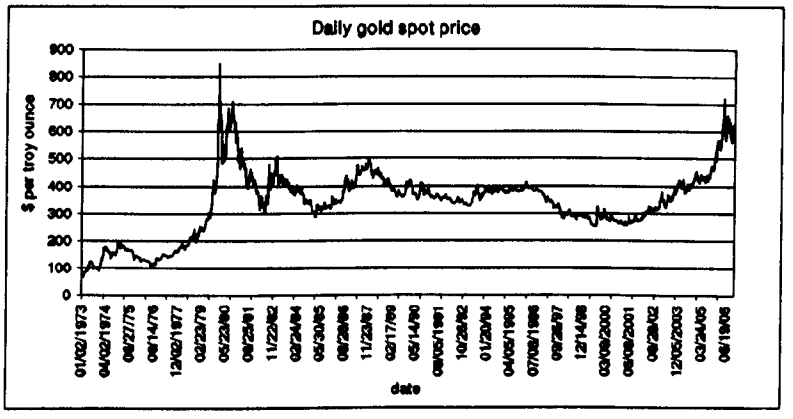
Table 3.1 exhibits the summary statistics of the gold price data set. The data ranges from a minimum of 63.9 dollars per troy ounce to a maximum of 850 dollars per troy ounce with a mean of 338.58 dollars per troy ounce.

daily gold prices (1973 - 2006)	
Mean	338.58
Standard Error	1.29172
Median	353.75
Mode	383
Standard Deviation	118.529261
Sample Variance	14049.1858
Kurtosis	0.481
Skewness	0.05387
Range	786.1
Minimum	63.9
Maximum	850
Count	8420
Confidence Level(95.0%)	2.5320953

Table 3.1: Summary statistics for daily gold prices

The daily gold spot prices are plotted in Figure 3.1. The gold spot prices run through different regimes. A possible segregation based on mean and standard deviation combinations is depicted below the graph. We emphasise that this choice

of different regimes is just one possibility; it shows that the characteristics of this data set vary through time.



Possible state segregation:			
time interval	mean	standard deviation	state*
01/02/73 - 12/29/78	147.46	93.17	3
01/02/79 - 12/30/82	438.97	129.43	1
01/04/83 - 12/31/98	372.68	43.82	2
01/04/99 - 12/31/02	284.57	19.6	3
01/02/03 - 11/23/06	452.35	92.62	1

* 1: high mean/ high std. dev. 2: high mean/ low std. dev. 3: low mean/ low std. dev.

Figure 3.1: Daily gold prices with the possible state segregation

We assume in this implementation that the log return's drift and volatility $\mathbf{f}$ and $\boldsymbol{\sigma}$, respectively, are governed by a Markov chain $\mathbf{x}$ with N states. Again, $\mathbf{x}$ is not directly observable but we do observe logarithmic increments y_{k+1} of the gold price process S_k. That is,

$$y_{k+1} = \ln \frac{S_{k+1}}{S_k} = \boldsymbol{f}_k + \boldsymbol{\sigma}_k w_{k+1} = \langle \mathbf{f}, \mathbf{x}_k \rangle + \langle \boldsymbol{\sigma}, \mathbf{x}_k \rangle w_{k+1}, \tag{3.14}$$

where w_k are standard Gaussian random variables. The implementation procedure starts by setting up initial values f_i and σ_i, $i = 1, 2, ..., N$ and a transition matrix (π_{ji}). These initial values are chosen based on the mean and variance of the data series. Since we have recursive formulae for the parameter estimates, the parameters are updated using the updated filtered values obtained from theorem 3.2. For this implementation, the filters are updated after every ten data points. Therefore, after ten values of y have been revealed, new estimates for $\mathbf{f}$, $\boldsymbol{\sigma}$ and Π are recursively computed. The EM algorithm guarantees the improvement of the model's parameters' estimates, the iterative estimates converge to local maxima

and because of this the model is 'self-tuning'.

Figures 3.2, 3.3 and 3.4 show the evolution of estimates for $\mathbf{f}$, $\boldsymbol{\sigma}$ and the transition matrix $\boldsymbol{\Pi}$ for a Markov chain with two, three and four states, respectively. In all three models parameters converge rapidly to a unique value regardless of the choice of initial estimates. The self-tuning algorithm reaches appropriate parameter values after approximately twelve passes. Since the EM algorithm only finds the local maxima, it is necessary to test the convergence of the algorithm for different choices of the initial values. We verified that this convergence can be achieved by different initial values. The evolution of estimates follows a similar pattern within the three-state model settings. The recursive formulae lead quickly to reasonable parameter estimates, which can be used in turn to forecast daily gold prices.

With the parameters estimated by the recursive formulae of the model, the h-step ahead predictions of the gold data time series can be calculated. From the semi-martingale representation (3.3) of $\mathbf{x}_k$, note that

$$\mathrm{E}[\mathbf{x}_{k+1}|\mathcal{F}_k^y] = \boldsymbol{\Pi}\widehat{\mathbf{x}}_k, \text{ where } \widehat{\mathbf{x}}_k = \mathrm{E}[\mathbf{x}_k|\mathcal{F}_k^y].$$

Consequently, from (2.5)

$$\mathrm{E}[\mathbf{x}_{k+h}|\mathcal{F}_k^y] = \boldsymbol{\Pi}^h\widehat{\mathbf{x}}_k \text{ for } h > 0. \tag{3.15}$$

From (3.15), the expectation and variance of the logarithmic increment y_{k+h}, which is a generalised version of (3.14), are

$$\mathrm{E}[y_{k+h}|\mathcal{F}_k^y] = \langle \mathbf{f}, \boldsymbol{\Pi}^{h-1}\widehat{\mathbf{x}}_k \rangle \tag{3.16}$$

and

$$\mathrm{Var}[y_{k+h}|\mathcal{F}_k^y] = \mathbf{f}^\top \mathrm{diag}(\boldsymbol{\Pi}^{h-1}\widehat{\mathbf{x}}_k)\mathbf{f} + \boldsymbol{\sigma}^\top \mathrm{diag}(\boldsymbol{\Pi}^{h-1}\widehat{\mathbf{x}}_k)\boldsymbol{\sigma} - \langle \mathbf{f}, \boldsymbol{\Pi}^{h-1}\widehat{\mathbf{x}}_k \rangle^2, \tag{3.17}$$

respectively, where $\mathrm{diag}(\boldsymbol{\Pi}^{h-1}\widehat{\mathbf{x}}_k)$ refers to the diagonal matrix whose diagonal entries are the components of the vector $\boldsymbol{\Pi}^{h-1}\widehat{\mathbf{x}}_k$.

To obtain the h-step ahead predictions of gold prices, for $h = 1, 2, 3, ..., 40$ we note that y_k is a random variable with a distribution which is a mixture of Gaussians. From (3.2) it follows that conditional on $\mathcal{F}_k^y$ the observation process y_{k+h} has

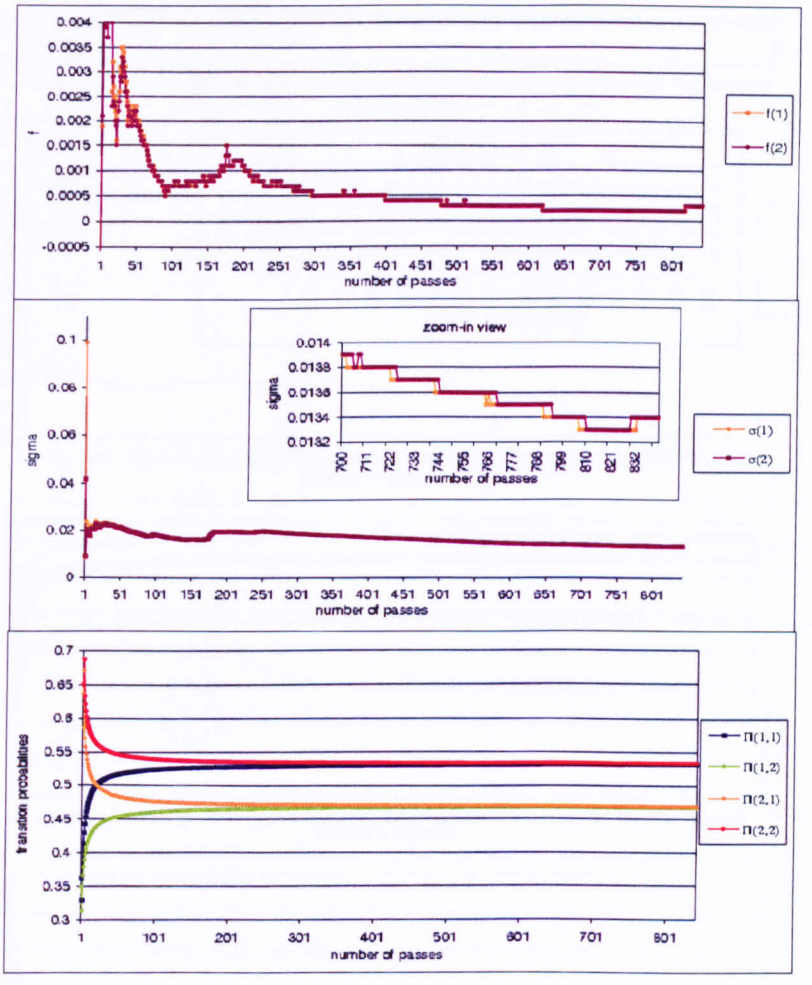


Figure 3.2: Evolution of estimates for $\boldsymbol{f}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\Pi}$-matrix ($N = 2$)

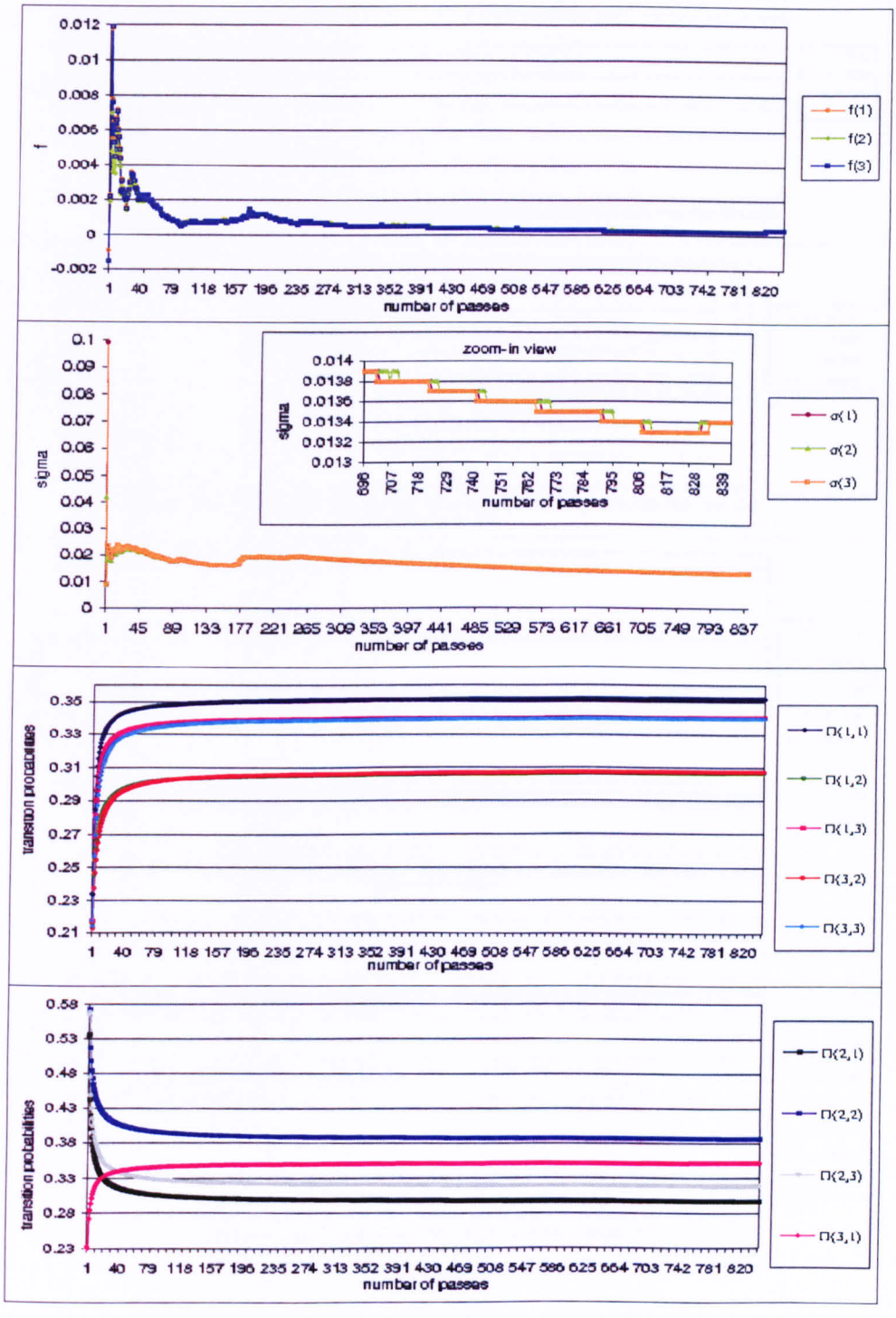


Figure 3.3: Evolution of estimates for $\boldsymbol{f}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\Pi}$-matrix ($N = 3$)

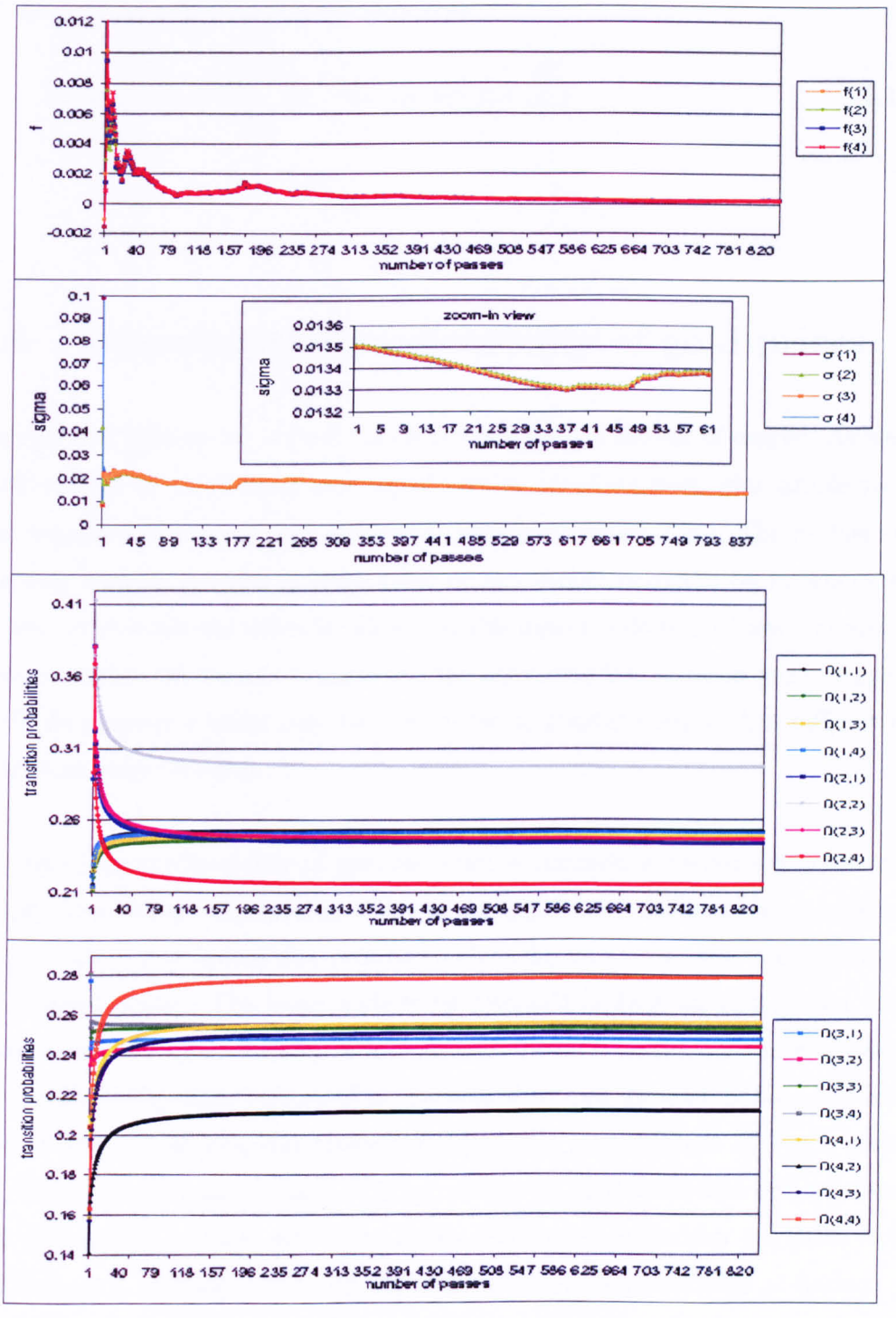


Figure 3.4: Evolution of estimates for $\boldsymbol{f}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\Pi}$-matrix ($N = 4$)

the distribution function $\sum_{i=1}^{N}\langle\hat{\mathbf{x}}_{k+h-1}, e_i\rangle\phi(y; f_i, \sigma_i^2)$. With (3.15) the expectation of the gold price at time $k+h$ is therefore

$$\mathrm{E}[S_{k+h}|\mathcal{F}_k^y] = S_k \sum_{i=1}^{N}\langle\mathbf{\Pi}^{h-1}\hat{\mathbf{x}}_k, e_i\rangle \exp\big(f_i + \frac{\sigma_i^2}{2}\big)\,. \tag{3.18}$$

3.5 Assessing the predictability of gold prices

Diebold and Killian [44] argued that predictability is a matter of degree. As well, predictability is a population property of a series, not of any particular sample path. The sample paths though can be used to estimate predictability. The problem of knowing whether a series is predictable or not should therefore be reformulated to *how* predictable the series is. However, this question does not have a clear-cut answer. A relevant forecast horizon and the associated loss function must be specified. Apparently a series may be predictable at shorter horizons but difficult to predict at longer horizons.

In comparing predictability of general series a common numéraire is necessary. Simply comparing the expected losses of forecasts for 2 series to judge relative predictability may ignore the possibility that the two series may be measured on different scales. The basic underlying principle of Diebold and Killian proposed measure of predictability is to base such measure on the difference between the conditionally expected loss of an optimal short-run forecast $E[L(\epsilon_{k+j,k})]$ and that of an optimal long-run forecast $E[L(\epsilon_{k+l,k})]$, $j \ll l$. Here $L(\cdot)$ is a given loss function and $\epsilon_{k+j,k} = y_{k+j} - \hat{y}_{k+j,k}$, where $\hat{y}_{k+j,k}$ denotes the optimal forecast of y_{k+j}. The series is said to be highly predictable at horizon j relative to l, if $E[L(\epsilon_{k+j,k})] \ll E[L(\epsilon_{k+l,k})]$ and the series is nearly unpredictable at horizon j relative to l if $E[L(\epsilon_{k+j,k})] \approx E[L(\epsilon_{k+l,k})]$. Consequently, this leads to the Diebold-Killian (DK) predictability measure

$$DK(L, \mathcal{F}_k, j, l) = 1 - \frac{E[L(\epsilon_{k+j,k})]}{E[L(\epsilon_{k+l,k})]}. \tag{3.19}$$

The usual loss functions employed are the symmetric, absolute value, quadratic and quartic types. Although in general, the only restrictions that $L(\cdot)$ must satisfy are (i) $L(0) = 0$ and (ii) $L(\cdot)$ is strictly monotone on each side of the origin. These ensure that the predictability measure in (3.19) with higher values indicates greater predictability.

The calculations of the DK predictability measures are based on the forecasts generated for daily gold prices between 1973 and 2006, a data set containing 8420 data points. The conditionally expected loss of an optimal short-run forecast $E[L(\epsilon_{k+j,k})]$ is calculated with two different loss functions L of 1-step ahead up to 20-step ahead forecasts ($j = 1, ...20$) at time k ($k = 1, ..., 8420$). The denominator of the DK predictability measure, namely the conditionally expected loss of an optimal long-run forecast $E[L(\epsilon_{k+l,k})]$ is calculated for $l = 10, l = 20, l = 30$ and $l = 40$. Therefore, the DK predictability measure is stated for four different numéraires l.

Figure 3.5 shows the plots of the DK measure of predictability given by (3.19) under two different loss functions (quadratic and quartic polynomials) applied to the forecast values calculated under 2-state, 3-state and 4-state hidden Markov chains. The DK predictability measure for four different numéraires ($l = 10, l = 20, l = 30, l = 40$) is plotted against the short-run forecast horizon $j = 1, ..., 20$. The higher the value of the DK metric, the greater is the degree of predictability.

The DK metric for 2-state, 3-state and 4-state Markov chain models show similar predictabilities with respect to each of the two loss functions. The predictability measures generated under the 4th degree polynomial loss function are the highest, followed by those obtained using the quadratic loss function. The differences of DK metric values between 2-, 3- and 4-state Markov chain models are close to zero. Apparently short-term forecasts for gold prices have a high degree of predictability under a 2-state as well as in a 3- or 4-state Markov chain model. In addition, the DK metric shows that for all three types of Markov chain models, highest degree of predictability is achieved for one-step ahead forecasts as expected. In particular, both the quadratic loss function and the 4th degree polynomial loss function

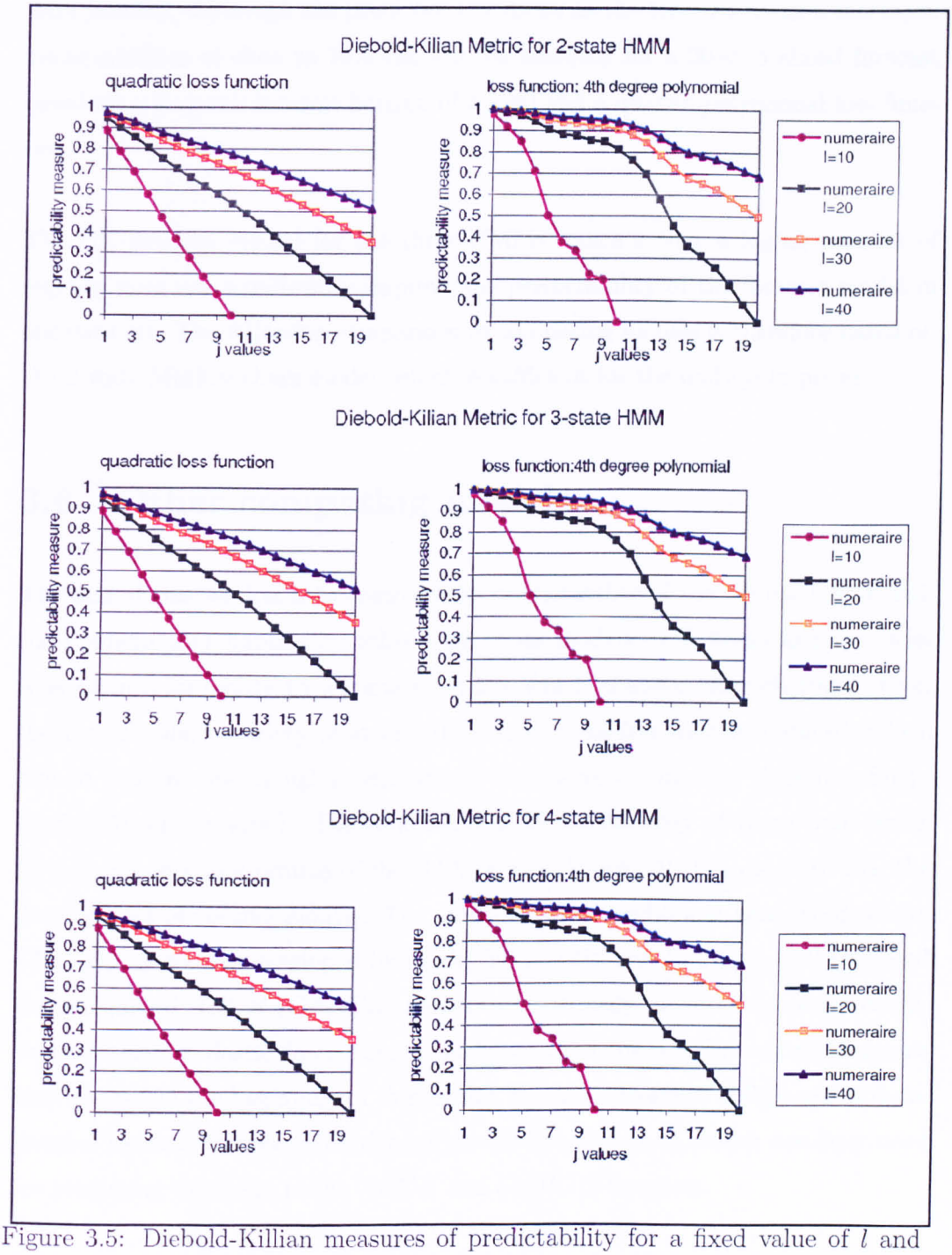


Figure 3.5: Diebold-Killian measures of predictability for a fixed value of l and varying j

indicate a predictability measure close to one for each of the numeraires for a short term horizon. Although the predictability drops as the forecast horizon increases, predictabilities of close to 70% can still be achieved for a 20-step ahead forecast based on a long-run forecast horizon of $l = 40$ and a quartic polynomial loss function.

The DK-measure results for the three HMMs indicate that a higher number of regimes does not significantly improve the predictability of the forecast model in our data set. The following comparison to competing models is therefore based on the 2-state Markov chain model, which is sufficient for the daily gold prices.

3.6 Other competing models

The aim of this section is to compare the predictability of the 2-state HMM with the predictability implied by other time series models. For financial time series it is often appropriate to assume volatility which changes through time. Over the last decade, a variety of stochastic volatility models were introduced, which aim to capture the changing variance or covariance of financial data (see Shephard [137] for a review). The conditional heteroscedasticity of these time series, for instance monthly returns of the DAX (see Lütkepohl [107]) have motivated the development of specific models. The autoregressive conditional heteroscedasticity ($ARCH$) model was developed by Engle [59] in 1982 and Bollerslev [24] proposed the generalised $ARCH$ ($GARCH$) model to realistically model and forecast conditionally heteroscedastic data. These models are widely used in forecasting economic time series. At the beginning of this section the autoregressive $AR(p)$ and autoregressive moving average $ARMA(p, q)$ models are described, which are later used for predicting the mean in the $ARCH$ and $GARCH$ forecasts.

Autoregressive (AR) models

An autoregressive process of order p, denoted $AR(p)$, satisfies

$$A_t = c + \Phi_1 A_{t-1} + \Phi_2 A_{t-2} + ... + \Phi_p A_{t-p} + \varepsilon_t \tag{3.20}$$

where $\{\varepsilon_t\}$ is a white noise sequence. The elements ε_t have mean zero and variance σ^2 and are IID. The order of the AR process is the longest time lag p associated with an A-term on the right-hand-side of (3.20), namely A_{t-p}. The coefficients $\Phi_1, ..., \Phi_p$ are constant terms, c is a constant related to the mean of the process. The process A_t is therefore modelled through p past values of the time series, $A_{t-1}, ..., A_{t-p}$, plus an error term ε_t.

Autoregressive Moving Average(ARMA) models

A generalisation of AR models are $ARMA$ models, which include moving average terms as well as autoregressive terms. An $ARMA$ process of order p and q, $ARMA(p, q)$, satisfies

$$\begin{aligned} A_t &= c + \Phi_1 A_{t-1} + \Phi_2 A_{t-2} + ... + \Phi_p A_{t-p} \\ &\quad + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon t - 2 + ... + \theta_q \varepsilon_{t-q} \end{aligned}$$

where $\{\varepsilon_t\}$ is a white noise sequence. The parameters $\Phi_1, ..., \Phi_p$ and $\theta_1, ..., \theta_p$ are constant. The orders of the $ARMA$ process, p and q, are the longest time lags, which are attached to an $A-$term and an error term, respectively.

Autoregressive Integrated Moving Average(ARIMA) models

Time series modelled by an $AR(p)$, $MA(q)$ or $ARMA(p, q)$ model are assumed to be stationary, that is the mean and the autocovariances of the process do not depend on the time t. If this is not the case, the time series can be differenced to achieve a stationary mean. The number of times the original series is differenced is denoted by d. Together with the order of the autoregressive part of the model p and the order of the moving average terms q it characterises the $ARIMA(p, d, q)$ model. The $d-$th difference of the time series is denoted by $\Delta_d A_t$. The process follows the equation

$$\begin{aligned} \Delta_d A_t &= c + \Phi_1 \Delta_d A_{t-1} + \Phi_2 \Delta_d A_{t-2} + ... + \Phi_p \Delta_d A_{t-p} \\ &\quad \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon t - q \end{aligned}$$

where $\{\varepsilon_t\}$, Φ_t and θ_t are defined as before.

Autoregressive conditional heteroscedasticity (ARCH) models

Let y_t denote the returns of a time series. The process y_t follows an autoregressive conditionally heteroscedastic process of order q ($ARCH(q)$), if

$$y_t = \sigma_t \epsilon_t \text{ with } \epsilon_t \sim N(0,1) \text{ and}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + ... + \alpha_q y_{t-q}^2, \; t = 1, ..., T$$

Given $Y_{t-1} := \{y_{t-1}, y_{t-2}, ...\}$, we have $y_t \mid Y_{t-1} \sim N(0, \sigma_t^2)$.

Generalised autoregressive conditional heteroscedasticity (GARCH) models

Bollerslev [24] proposed in 1986 an extension of the ARCH models, the generalised autoregressive conditional heteroscedasticity models of order p and q ($GARCH(p,q)$). The conditional variance in this model is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + ... + \alpha_p y_{t-p}^2 + \imath_1 \sigma_{t-1}^2 + ... + \imath_q \sigma_{t-q}^2$$

3.6.1 Comparison of forecasts

In this section the predictability of the 2-state hidden Markov model is compared to the predictability implied by other widely used time series models. The first part shows the results of $ARMA$ models, which were defined above. In the second part we compare our model to $ARCH$ and $GARCH$ models.

The models AR, $ARMA$ and $ARIMA$ are implemented with the System Identification Toolbox of MATLAB. The optimal lag order for the AR models is determined using the Akaike Information Criterion (AIC). This method was introduced by Akaike in 1973 [1]. When comparing several models with different specification of the probability density function $f(x|\theta)$, where θ is a vector parameter, the one chosen has the Minimum Information Criterion Estimate (MAICE), defined by $f(x|\hat{\theta})$, which gives the minimum AIC.

The optimal lag order for the AR models is determined using the AIC function in MATLAB. Here, the AIC is calculated by

$$AIC = \log(Loss) + \frac{2no_p}{DV}$$

where *Loss* is the loss function, no_p is the number of parameters and DV is the number of estimated data values. The loss function is defined by

$$Loss = \det(\frac{1}{DV}\sum_{i=1}^{DV}\epsilon(t,\hat{\theta}_{DV})(\epsilon(t,\hat{\theta}_{DV}))^T).$$

The parameter estimate is denoted by $\hat{\theta}_{DV}$. For a derivation of this formula from the definition in Akaike [1], see Ljung [102]. Within the data set of daily gold spot price, the optimal lag order leads to implementation of the models AR(4), ARMA(5,2) and ARIMA(2,1,2).

By the prediction error/maximum likelihood method (see Ljung [102] for details) we calculate the time series model parameters and obtain the following:

$$\begin{aligned}
AR(4): \quad & y_t = 0.9517y_{t-1} + 0.03183y_{t-2} + 0.08458y_{t-3} - 0.06805y_{t-4} + \varepsilon_t \\
ARMA(5,2): \quad & y_t = 2.05y_{t-1} - 1.749y_{t-2} + 0.6629y_{t-3} + 0.004688y_{t-4} + 0.03157y_{t-5} \\
& \quad + \varepsilon_t + 1.148\varepsilon_{t-1} - 0.7967\varepsilon_{t-2} \\
ARIMA(2,1,2): \quad & \triangle y_t = 0.3431\triangle y_{t-1} - 0.308\triangle y_{t-2} + \varepsilon_t - 0.438\varepsilon_{t-1} \\
& \quad - 0.3321\varepsilon_{t-2}
\end{aligned}$$

where $\triangle$ denotes the first forward difference operator and ε is the error term. Then the DK-measure of predictability is examined for the three models using the forecasts generated by the above models. Again, two different loss functions are used in the analysis with a long-run forecast horizon l as the numéraire and the near-term forecast horizon j.

Figures 3.6 and 3.7 exhibit the DK measures for the *AR*, *ARMA* and *ARIMA* models, respectively. It has to be noted that the predictability of the short-term forecasts is significantly higher than that of the long-term forecasts in all three models.

Table 3.2 shows the DK-measures for the 2-state HMM and the three time series models for 1- and 20-step ahead predictions with the numéraire $l = 40$. The DK-measure of $ARIMA(2,1,2)$ examined with the quadratic loss function for a 1-step ahead forecast is slightly but not significantly higher than that of the 2-state HMM.

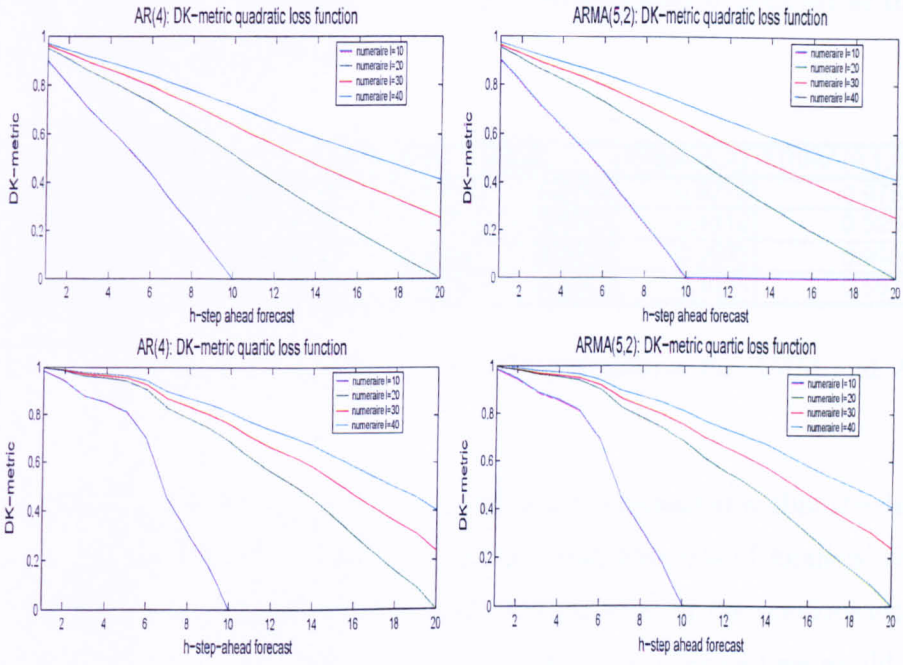


Figure 3.6: Diebold-Kilian measure for $AR(4)$ and $ARMA(5,2)$

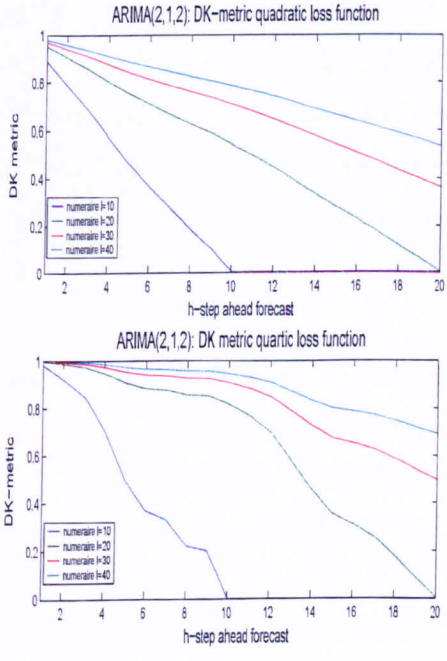


Figure 3.7: Diebold-Kilian measure for $ARIMA(2,1,2)$

However,the 2-state HMM indicates a higher predictability of gold prices than the $AR(4)$ and $ARMA(5,2)$ model.

		2-state MC	AR(4)	ARMA(5,2)	ARIMA(2,1,2)
Quadratic loss function	DK(1,40)	0.9750	0.9724	0.9725	0.9752
	DK(20,40)	0.5142	0.4113	0.4110	0.5264
Quartic loss function	DK(1,40)	0.9989	0.9973	0.9973	0.9989
	DK(20,40)	0.6858	0.4010	0.4025	0.6876

Table 3.2: Comparison of the DK predictability measures for HMM and $ARMA$ models

In addition, the $ARCH$ and $GARCH$ models are implemented within the data set, since they are widely used to generate h-step ahead forecasts of financial time series, mostly for modelling stochastic volatility. Motivated by the previous empirical research on $ARCH$ and $GARCH$ (see Taylor [143] for a review) we would like to forecast the daily gold spot prices with an $ARCH(1)$ and $GARCH(1,1)$ model and compare their forecastability performance to that of the HMM. First the data set is further analysed to determine wether these models are applicable to this data set. The first graph in Figure 3.8 depicts the daily returns of the daily spot price data from 1973 until 2006. The second graph in Figure 3.8 shows the autocorrelation of the returns. With T denoting the length of the return series, the autocorrelation are within the $\pm\frac{2}{\sqrt{T}}$ band and are therefore uncorrelated. However, depicting the autocorrelation of the squared returns shows autocorrelation, therefore the data set reveals the characteristic of heteroscedasticity.

In recognition of the presence of heteroscedasticity revealed by the data, the predictability of the 2-state and 3-state hidden Markov model is compared to the predictability implied by the $ARCH(1)$ and $GARCH(1,1)$ models. The DK measure of predictability is examined for the $ARCH(1)$ and $GARCH(1,1)$ using the forecasts obtained from these models. The $ARCH/GARCH$ calculations were performed using the MATLAB GARCH toolbox [113]. The function "garchpred" computes forecasts for the conditional mean and standard deviation of an observed univariate return series based on the $ARMA$ and $GARCH$ model estimated with the function "garchfit". The forecast is performed within "rolling windows"; for every data set an $ARCH/$ $GARCH$ model is determined and one- to fourty-step

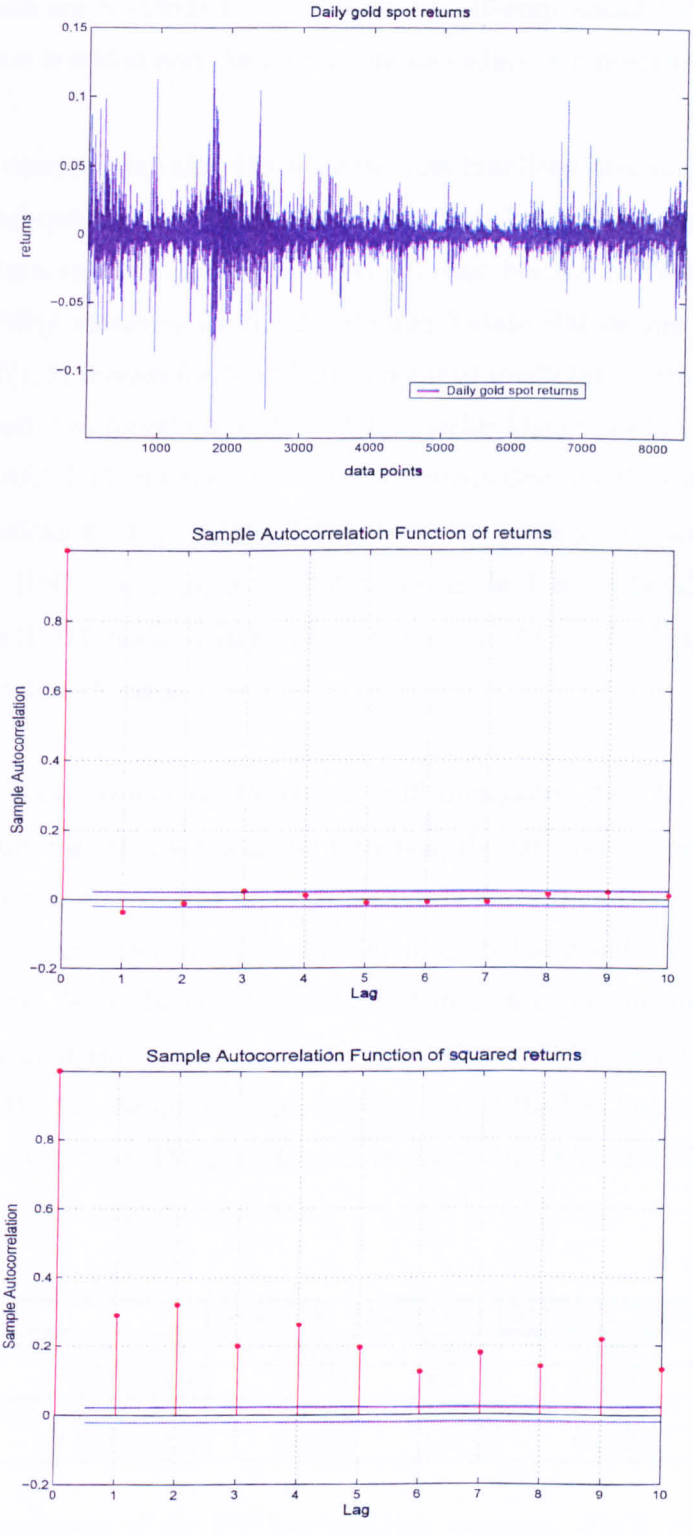


Figure 3.8: Plot of daily returns and sample autocorrelation values for returns of daily gold prices

ahead forecasts are performed. After each 1-to 40-step ahead forecast one data point at a time is added and the forecasting procedure is repeated.

To make the comparison valid, the same two loss functions described in section 3.5 (quadratic and quartic) are utilised in this analysis here with a long-run forecast horizon l as the numéraire and the near-term forecast horizon j. Table 3.3 shows the DK predictability measures for the 2-state and 3-state HMMs and the $ARCH(1)$ and $GARCH(1,1)$ models for 1- and 20-step ahead predictions with the numéraire $l = 40$. For both loss functions, $GARCH(1,1)$ yields higher predictability measure than that of $ARCH(1)$ for the one-step ahead prediction but the difference has no practical significance. The 3-state HMM shows the same predictability with that of the 2-state HMM. It is apparent that as far as the DK predictability metric is concerned the HMM model performs better than the $ARCH/GARCH$ models for this very short time horizon.

On the other hand, whilst $GARCH(1,1)$ still dominates $ARCH(1)$ in terms of the DK measure for the 20-step ahead forecasts, the DK measure generated from the HMM with the quartic loss function is slightly lower than those of the $ARCH$ and $GARCH$ models. Nevertheless, the difference is too small to yield any practical significance. Note though that the HMM gives a higher predictability value than the other models for short- and long-term predictability under the quadratic loss function. Within the market data considered and the loss functions employed, these findings show that HMM tends to outperform the $ARCH/GARCH$ models in terms of the short-run forecasts only.

		3-state MC	2-state MC	GARCH(1,1)	ARCH(1)
Quadratic loss function	DK(1,40)	0.9750	0.9750	0.9287	0.9237
	DK(20,40)	0.5142	0.5142	0.4833	0.4793
Quartic loss function	DK(1,40)	0.9989	0.9989	0.9871	0.9885
	DK(20,40)	0.6858	0.6858	0.6892	0.6712

Table 3.3: Comparison of the DK predictability measure: HMM versus $ARCH/GARCH$ models

3.6.2 Measuring forecast errors

Additionally, we also evaluate the forecasting errors for the three different types of models. We first define three commonly used error measures, namely the median relative absolute error (MdRAE), median absolute percentage error (MdAPE) and mean square error (MSE). Armstrong and Collopy [4] investigated the reliability of different error measure and recommend MdRAE and MdAPE. The MSE is not a unit-free error measure, but since it is widely used by practitioners, this measure is included in the error analysis. Let $A_{h,series}$ be the actual value at horizon h of series s and $F_{met,h,series}$ be the forecast from method *met* for the forecast horizon h of series *series* with $OB = number\,of\,observations$. The relative absolute error sets the forecast from method *met* in relation to the forecast by the random walk and this is denoted by $F_{rw,h,series}$.

1. Median relative absolute error (MdRAE)

$$RAE_{met,h,series} = \left| \frac{F_{met,h,series} - A_{h,series}}{F_{rw,h,series} - A_{h,series}} \right|$$

   Median RAE: Observation $\frac{OB+1}{2}$ if OB is odd or the mean of observation $\frac{OB}{2}$ and $\frac{OB}{2} + 1$ if OB is even (observations are rank-ordered by RAE).

2. Median absolute percentage error

$$APE_{met,h,series} = \left| \frac{F_{met,h,series} - A_{h,series}}{A_{h,series}} \right|$$

   Median APE: Observation $\frac{OB+1}{2}$ if S is odd or the mean of observation $\frac{OB}{2}$ and $\frac{OB}{2} + 1$ if S is even (observations are rank-ordered by APE).

3. Mean square error

$$\frac{1}{OB} \sum_{series=1}^{OB} (F_{met,h,series} - A_{h,series})^2 .$$

These error measures are used to further evaluate the h-step ahead forecasts ($h = 1, 2, \ldots, 20$). The forecasting errors are reported in Table 3.4. In the HMM column of Table 3.4, no size of the state is indicated; this is because from our calculations

the 2- and 3-state HMM models produce the same forecasting errors for all forecast horizons. This result reinforces further the adequacy of a 2-state HMM in capturing the dynamics of this data set. Comparing the forecasting errors of the HMM with those of $GARCH$ and $ARCH$ models, we see that the 1-step ahead forecasts in the HMM setting are closest to the actual data. However, an error comparison of the 10- and 20-step ahead forecasts, for example, shows smaller forecasting errors for $ARCH(1)$ and $GARCH(1,1)$. But still, it has to be noted that based on the short-term horizon forecast obtained through the HMM method, HMM clearly outperforms the $ARCH(1)$ and $GARCH(1,1)$ models.

	MdRAE			MdAPE			MSE		
h-step ahead	HMM	GARCH (1,1)	ARCH(1)	HMM	GARCH (1,1)	ARCH(1)	HMM	GARCH (1,1)	ARCH(1)
1	0.5067	1.0083	1.4330	0.0055	0.0108	0.0142	30.2388	93.8189	68.8049
2	0.6619	1.0272	1.6018	0.0076	0.0110	0.0162	57.4875	94.4108	71.5067
3	0.8191	1.0466	1.6760	0.0094	0.0112	0.0170	83.6564	94.9900	72.7552
4	0.9537	1.0629	1.7083	0.0110	0.0113	0.0174	114.0911	95.5590	73.3212
5	1.0924	1.0755	1.7218	0.0128	0.0115	0.0176	144.4212	96.1198	73.5758
6	1.2209	1.0881	1.7301	0.0141	0.0116	0.0177	171.5325	96.6737	73.6931
7	1.3192	1.1005	1.7367	0.0155	0.0118	0.0178	196.9294	97.2220	73.7511
8	1.4455	1.1132	1.7382	0.0165	0.0119	0.0178	221.3671	97.7655	73.7834
9	1.5338	1.1221	1.7406	0.0179	0.0120	0.0178	244.7163	98.3048	73.8041
10	1.6354	1.1328	1.7440	0.0188	0.0122	0.0178	272.4336	98.8404	73.8195
11	1.7165	1.1466	1.7444	0.0198	0.0123	0.0178	299.7890	99.3728	73.8321
12	1.8087	1.1569	1.7446	0.0210	0.0124	0.0178	327.8202	99.9022	73.8431
13	1.8940	1.1709	1.7458	0.0220	0.0125	0.0178	361.3668	100.4289	73.8531
14	1.9686	1.1856	1.7461	0.0226	0.0126	0.0178	393.3900	100.9531	73.8624
15	2.0780	1.1955	1.7468	0.0236	0.0127	0.0178	424.5577	101.4750	73.8712
16	2.1451	1.2092	1.7468	0.0243	0.0129	0.0178	454.9274	101.9947	73.8795
17	2.2048	1.2195	1.7468	0.0248	0.0130	0.0178	486.2919	102.5122	73.8874
18	2.2697	1.2297	1.7468	0.0256	0.0131	0.0178	519.8710	103.0278	73.8950
19	2.3391	1.2431	1.7470	0.0264	0.0132	0.0178	552.2800	103.5414	73.9023
20	2.4120	1.2526	1.7470	0.0270	0.0134	0.0178	586.4499	104.0532	73.9094

Table 3.4: Error analysis

A plot comparing 1-step ahead forecasts generated from the 2-state HMM, $ARCH(1)$ and $GARCH(1,1)$ is depicted in Figure 3.9. This zoom-in view over a one year span of the considered time period shows clearly how close the 1-step ahead forecasts follow the actual data. On the basis of the previous error analysis the set of 1-step ahead forecasts generated with the 2-state HMM shows the least deviation from the actual daily gold prices.

The MATLAB code for the implementation of the filters and optimal parameter estimates as well as the calculation of the DK metric can be found in Appendix

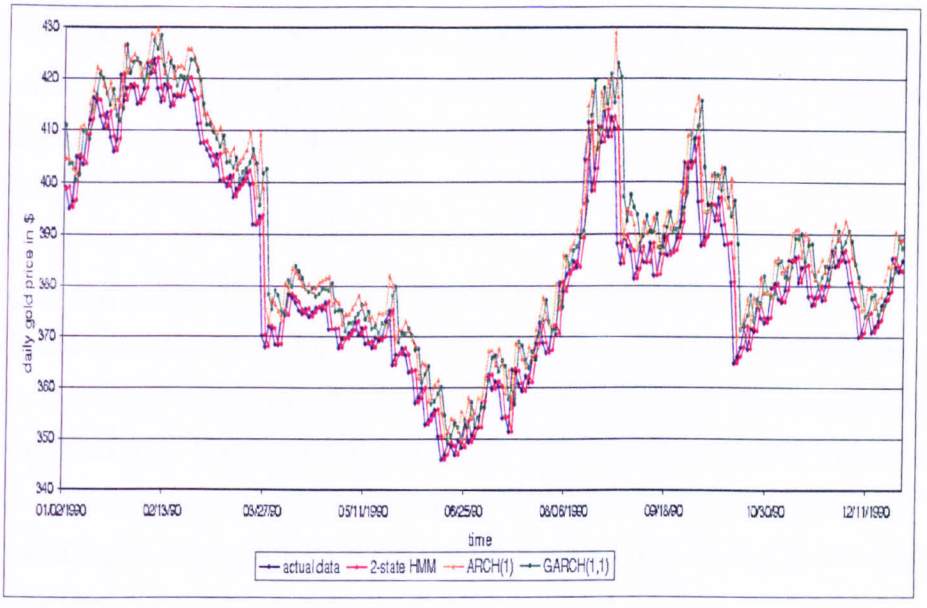


Figure 3.9: Comparison of 1-step ahead forecasts

B.3.

3.7 Some concluding remarks

In this chapter a filtering technique from Elliott [49] was adopted but refined in the study of the commodity markets. In particular, the use of HMM filtering theory is considered to investigate the forecastability of daily spot prices in the gold market. Optimal recursive estimation procedures for the model parameters were presented. The model parameters switch between regimes and the calibrated parameter estimates are governed by an $N-$ state Markov chain. The filtering technique makes use of the EM algorithm in conjunction with the change of probability measure technique. The forecasts for h-step ahead predictions is analysed via the DK metric, which provides measures of forecastability. As expected, the DK metric decreases significantly as forecast horizon increases. A high degree of forecastability is attained for one-step ahead predictions, whereby both loss functions, independent of the chosen number of states for the Markov chain, give a predictability of close to 98%. A comparison of the forecastability of HMMs to

other time series models, namely $AR(4)$, $ARMA(5,2)$ and $ARIMA(2,1,2)$ as well as to $ARCH(1)$ and $GARCH(1,1)$ is conducted. The 2-state and 3-state HMMs produce higher measures of short- and medium-run forecastability in the empirical application.

The question of how to calculate optimally the number of states implied by the data is not considered here but this issue will be re-visited and addressed in chapter 5. The result in this chapter however is consistent with previous findings of other authors (e.g. Hamilton [79] and Hardy [85]) wherein a 2- or 3-state Markov chain is sufficient and reasonable to capture the stylised facts of market data. The number of states was increased in the implementation, but no significant improvements in terms of goodness of fit are further achieved.

Chapter 4

A general filtering technique

In this chapter, the general filtering techniques and the filter equations which were established by Elliott [48] for Markov chains observed in Gaussian noise are described. These filters are used in chapters 5 and 6 for two different modelling frameworks to filter out information about the underlying Markov chain from the observation process. In general, we derive filters for four types of processes related to the Markov chain, namely the state space process, the jump process, the occupation time process and auxiliary processes including terms of the observation process. Throughout this work we will have the following framework: let $(\Omega, \mathcal{F}, P)$ be the underlying probability space of a homogeneous Markov chain $\mathbf{x}_k$ with finite state in discrete time ($k = 0, 1, ...$). The distribution of $\mathbf{x}_0$ is known and the state space of $\mathbf{x}_k$ is associated with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ of $\mathbb{R}^N$. The i^{th} vector $\mathbf{e}_i$ is given by $\mathbf{e}_i = (0, ..., 1, ..., 0)^\top$; that is, $\mathbf{e}_i$ has 1 in its i^{th} component and 0 elsewhere. Let $\mathcal{F}_0^X = \sigma\{\mathbf{x}_0, \ldots, \mathbf{x}_k\}$ be the $\sigma-$field generated by $\mathbf{x}_0, \ldots, \mathbf{x}_k$, and $\mathcal{F}_k^X$ be the complete filtration generated by $\mathcal{F}_0^X$. Furthermore let $\mathcal{F}_k^y$ denote the complete filtration generated by the observation process y, so that $\mathcal{F}_k = \mathcal{F}_k^X \vee \mathcal{F}_k^y$ is the global filtration generated by $\mathbf{x}$ and y. Under the real world probability measure P, the Markov chain $\mathbf{x}$ has dynamics

$$\mathbf{x}_{k+1} = \mathbf{\Pi}\mathbf{x}_k + \mathbf{v}_{k+1} \tag{4.1}$$

where $\mathbf{v}_{k+1}$ is a martingale increment with $\mathrm{E}[\mathbf{v}_{k+1} \mid \mathcal{F}_k^X] = 0$, $\mathbf{\Pi} = (\pi_{ji})$ is the transition probability matrix and $\pi_{ji} = P(\mathbf{x}_{k+1} = \mathbf{e}_j | \mathbf{x}_k = \mathbf{e}_i)$.

The first section of this chapter describes a general filtering technique whilst the adaptive filters are derived in section 4.2.

4.1 Filtering technique

To determine the expectation of any $\mathcal{F}-$adapted stochastic process H given the filtration $\mathcal{F}_k^y$, consider the reference probability measure $\bar{P}$ defined as

$$P(A) = \int_A \bar{\Lambda}\, d\bar{P}. \tag{4.2}$$

From Bayes' theorem 2.3, a filter for any adapted process H is given by

$$\mathrm{E}\left[H_k \mid \mathcal{F}_k^y\right] \;=\; \frac{\bar{E}\left[H_k\bar{\Lambda}_k \mid \mathcal{F}_k^y\right]}{\bar{E}\left[\bar{\Lambda}_k \mid \mathcal{F}_k^y\right]} \quad .$$

We define $\eta_k(H_k) := \bar{\mathrm{E}}\left[H_k\bar{\Lambda}_k \mid \mathcal{F}_k^y\right]$, so that $\mathrm{E}\left[H_k \mid \mathcal{F}_k^y\right] = \frac{\eta_k(H_k)}{\eta_k(1)}$. A recursive relationship between $\eta_k(H_k)$ and $\eta_{k-1}(H_{k-1})$ has to be found, where $\eta_0(H_0) = E[H_0]$. However, we shall first find a recursive formula for the term $\eta_{k-1}(H_{k-1}\mathbf{x}_{k-1})$. Note that H_k is scalar whilst $\eta_{k-1}(H_{k-1}\mathbf{x}_{k-1})$ is a vector. To relate $\eta_k(H_k)$ and $\eta_k(H_k\mathbf{x}_k)$ we note further that with $\langle \mathbf{1}, \mathbf{x}_k\rangle = 1$

$$\langle \mathbf{1}, \eta_k(H_k\mathbf{x}_k)\rangle = \eta_k(H_k\langle \mathbf{1}, \mathbf{x}_k\rangle) = \eta_k(H_k). \tag{4.3}$$

Therefore

$$\mathrm{E}\left[H_k \mid \mathcal{F}_k^y\right] \;=\; \frac{\langle \mathbf{1}, \eta_k(H_k\mathbf{x}_k)\rangle}{\langle \mathbf{1}, \eta_k(\mathbf{x}_k)\rangle}.$$

Suppose H_l is a scalar $\mathcal{F}-$adapted process, H_0 is $\mathcal{F}_0^{\mathbf{x}}$ measurable and

$$H_l = H_{l-1} + a_l + \langle b_l, \mathbf{v}_l\rangle + g_l f(y_l)$$

where a, b and g are $\mathcal{F}$-predictable, f is a scalar-valued function and $\mathbf{v}_l = \mathbf{x}_l - \Pi\mathbf{x}_{l-1}$. A recursive relation for $\eta_k(H_k\mathbf{x}_k)$ is given by

$$\begin{aligned}
\eta_k(H_k\mathbf{x}_k) \;=\; \sum_{i=1}^{n} \Gamma^i(y_k)\big[&\langle \mathbf{e}_i, \eta_{k-1}(H_{k-1}\mathbf{x}_{k-1})\rangle \Pi\mathbf{e}_i \\
&+ \langle \mathbf{e}_i, \eta_{k-1}(a_k\mathbf{x}_{k-1})\rangle \Pi\mathbf{e}_i \\
&+ (\mathrm{diag}(\Pi\mathbf{e}_i) - (\Pi\mathbf{e}_i)\otimes(\Pi\mathbf{e}_i))\eta_{k-1}(b_k\langle \mathbf{e}_i, \mathbf{x}_{k-1}\rangle) \\
&+ \eta_{k-1}(g_k\langle \mathbf{e}_i, \mathbf{x}_{k-1}\rangle) f(y_k)\Pi\mathbf{e}_i\big]
\end{aligned} \tag{4.4}$$

Here, for any column vectors $\mathbf{z}$ and $\mathbf{y}$, $\mathbf{z} \otimes \mathbf{y}$ denotes the rank-one (if $\mathbf{z} \neq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$) matrix $\mathbf{z}\mathbf{y}^\top$. The proof of this formula can be found in Elliott [48], theorem 5.3. The term $\Gamma^i(y_k)$ denotes the component-wise Radon-Nikodým derivative $\bar{\lambda}_k^i$, which is constructed in accordance with the dynamics of the observation process y_k. This term therefore varies with each model setting and will have a form specific to each model.

4.2 Adaptive and recursive filters

For the estimation of the unknown parameters of an observation process, hidden information is filtered out from the observation process through an estimator for the state of the Markov chain as well as for three related processes: the jump process, the occupation time process and auxiliary processes of the Markov chain. These processes can be characterised as special cases of the general process $H_l = H_{l-1} + a_l + \langle b_l, \mathbf{v}_l \rangle + g_l f(y_l)$, where H_0 is $\mathcal{F}_0^X$ measurable.

The estimator for the state $\mathbf{x}_k$ in equation (4.4) is derived from $\eta_k(H_k \mathbf{x}_k)$ by setting $H_k = H_0 = 1$, $a_k = 0$, $b_k = 0$ and $g_k = 0$. This implies that

$$\eta_k(\mathbf{x}_k) = \sum_{i=1}^{N} \Gamma^i(y_k) \langle \mathbf{e}_i, \eta_{k-1}(\mathbf{x}_{k-1}) \rangle \Pi \mathbf{e}_i \quad . \tag{4.5}$$

The first related process is the number of jumps of the Markov chain $\mathbf{x}_k$ from state $\mathbf{e}_r$ to state $\mathbf{e}_s$ in time k :

$$\begin{aligned} J_k^{(sr)} &= \sum_{l=1}^{k} \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_l, \mathbf{e}_s \rangle \\ &= J_{k-1}^{(sr)} + \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \pi_{sr} + \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \langle \mathbf{v}_k, \mathbf{e}_s \rangle \quad . \end{aligned} \tag{4.6}$$

Setting $H_k = J_k^{(sr)}$, $H_0 = 0$, $a_k = \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \pi_{sr}$, $b_k = \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \mathbf{e}_s'$ and $g_k = 0$ in equation (4.4) we get

$$\begin{aligned}\eta_k(J_k^{sr}\mathbf{x}_k) &= \sum_{i=1}^{N}\Gamma^i(y_k)\overline{E}\big[\Lambda_{k-1}\langle\mathbf{x}_{k-1},\mathbf{e}_i\rangle\{J_{k-1}^{sr}\Pi\mathbf{e}_i+\langle X_{k-1},\mathbf{e}_r\rangle\pi_{sr}\Pi\mathbf{e}_i \\ &\quad +\langle\mathbf{x}_{k-1},\mathbf{e}_r\rangle\mathbf{e}_s'(\mathrm{diag}\Pi\mathbf{e}_i-(\Pi\mathbf{e}_i)\otimes(\Pi\mathbf{e}_i))\}\big] \\ &= \sum_{i=1}^{N}\Gamma^i(y_k)\langle\eta_{k-1}(J_{k-1}^{sr}\mathbf{x}_{k-1}),\mathbf{e}_i\rangle\Pi\mathbf{e}_i \\ &\quad +\Gamma^r(y_k)\eta_{k-1}(\langle\mathbf{x}_{k-1},\mathbf{e}_r\rangle)\pi_{sr}\mathbf{e}_s \quad . \end{aligned} \tag{4.7}$$

The second process $O_k^{(r)}$ denotes the occupation time of the Markov process $\mathbf{x}$, which is the length of time $\mathbf{x}$ spent in state r up to time k. Here,

$$O_k^r = \sum_{l=1}^{k}\langle\mathbf{x}_{l-1},\mathbf{e}_r\rangle = O_{k-1}^r+\langle\mathbf{x}_{k-1},\mathbf{e}_r\rangle\,. \tag{4.8}$$

We set $H_k = O_k^r$, $H_0 = 0$, $a_k = \langle\mathbf{x}_{k-1},\mathbf{e}_r\rangle$, $b_k = 0$ and $g_k = 0$ in equation (4.4) to obtain

$$\begin{aligned}\eta_k(O_k^r\mathbf{x}_k) &= \sum_{i=1}^{N}\Gamma^i(y_k)\{\langle\eta_{k-1}(O_{k-1}^r\mathbf{x}_{k-1}),\mathbf{e}_i\rangle\Pi\mathbf{e}_i \\ &\quad +\eta_{k-1}(\langle\mathbf{x}_{k-1},\mathbf{e}_r\rangle\langle\mathbf{x}_{k-1},\mathbf{e}_i\rangle)\Pi\mathbf{e}_i\} \\ &= \sum_{i=1}^{N}\Gamma^i(y_k)\langle\eta_{k-1}(O_{k-1}^r\mathbf{x}_{k-1},\mathbf{e}_i\rangle\Pi\mathbf{e}_i \\ &\quad +\Gamma^r(y_k)\langle\eta_{k-1}(\mathbf{x}_{k-1}),\mathbf{e}_r\rangle\Pi\mathbf{e}_r \quad . \end{aligned} \tag{4.9}$$

Finally, consider the auxiliary process $T_k^r(g)$, which occur in the maximum likelihood estimation of model parameters. Specifically,

$$\begin{aligned}T_k^{(r)}(g) &= \sum_{l=1}^{k}\langle\mathbf{x}_{l-1},\mathbf{e}_r\rangle g(y_l) \\ &= T_{k-1}^r(g)+\langle\mathbf{x}_{k-1},\mathbf{e}_r\rangle g(y_k)\end{aligned} \tag{4.10}$$

where g is a function of the form $g(y) = y$, $g(y) = y^2$, $g(y) = y_{l+1}$, $g(y) = y_{l+1}y_l$ or $g(y) = y_{l+1}^2, 1 \le l \le k$. We apply formula (4.4) with the substitution $H_k = T_k^r(g)$,

$H_0 = 0$, $a_k = 0$, $b_k = 0$ and $g_k = \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle$ and get

$$\begin{aligned}
\eta_k(T_k^r(g)\mathbf{x}_k) &= \sum_{i=1}^{N} \Gamma^i(y_k)\{\langle \eta_{k-1}(T_{k-1}^r(g)\mathbf{x}_{k-1}), \mathbf{e}_i\rangle \Pi \mathbf{e}_i \\
&\quad + \eta_{k-1}(\langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_i\rangle) g(y_k) \Pi \mathbf{e}_i\} \\
&= \sum_{i=1}^{N} \Gamma^i(y_k)\{\langle \eta_{k-1}(T_{k-1}^r(g)\mathbf{x}_{k-1}), \mathbf{e}_i\rangle \Pi \mathbf{e}_i \\
&\quad + \Gamma^r(y_k)\langle \eta_{k-1}(\mathbf{x}_{k-1}), \mathbf{e}_r\rangle g(y_k) \Pi \mathbf{e}_r \,. \qquad (4.11)
\end{aligned}$$

The recursive optimal estimates of J, O and T can be calculated using equation (4.3).

Chapter 5

A hidden Markov model for interest rates

The short term interest rate is a key variable in financial modelling, because of its significance in pricing and hedging of fixed income securities and other financial derivatives. In this chapter a hidden Markov model in discrete time for interest rates is developed and recursive formulae for the estimation of the parameters in this proposed model are derived. The first section in this chapter reviews some short term interest rate models with and without regime-switching. The framework for the hidden Markov model is described in section 5.2. In section 5.3, recursive parameter estimations are derived utilizing the filters for the Markov process. The model is implemented on 30-day Treasury-bill rates in section 5.4, a sensitivity analysis is obtained and a simple technique on how to choose the optimal number of states is discussed.

5.1 Short rate models

The instantaneous spot interest rate is the basis for modelling the term structure of interest rates. Several one-factor models for the short term interest rate r were developed in recent years, most of which are based on a general stochastic differential

equation

$$dr_t = \mu(t, r_t)dt + u(t, r_t)\, dW_t \tag{5.1}$$

where W_t is a standard Brownian motion under a probability measure P. Well-known models include a no-arbitrage single-factor model by Vasicek [145], where the interest rate is mean-reverting and its extension developed by Hull and White [91] having the form

$$dr_t = [\theta_t - a_t r_t]\, dt + u_t\, dW_t\,. \tag{5.2}$$

The parameters a_t, θ_t and u_t are deterministic functions of time t. The short-term interest rates can be fitted to today's term structure by choosing θ_t appropriately. The Wiener process W_t is independent of θ_t. Also, the volatility of the short rate at future times is described by u_t and a_t is the mean-reversion rate. One drawback of this model is that the interest rates are normally distributed, therefore they can become negative. On the other hand, the model is highly tractable and closed-form bond prices can be derived, which makes it very popular, see Pelsser [126].

Cox, Ingersoll and Ross [35] proposed a single-factor model, where positive interest rates are guaranteed. The short rate r has a noncentral chi-square distribution and it has the dynamics

$$dr_t = [\theta - a r_t]\, dt + u\sqrt{r_t}\, dW_t \tag{5.3}$$

for constants θ, a and u. The short rate is mean-reverting and it tends towards $\frac{\theta}{a}$. In the generalised version of the Cox-Ingersoll-Ross (CIR) model, the parameters θ, a and u are deterministic functions of time t. The model can then be fitted to the initial term structure.

Furthermore there are several multi-factor models for interest rates (e.g. Duffie and Kan [46] and Longstaff and Schwartz [103], amongst many others). These models are able to provide a better fit to the yield curve but the analysis and parameter estimation are more difficult. Another popular interest rate model is the lognormal short rate model by Black and Karasinski [21]. Heath, Jarrow and Morton [89] take a somewhat different methodology, where the entire forward-rate

curve is modelled. A detailed discussion of these models and methodologies can be found in Brigo and Mercurio [27].

5.1.1 Regime-switching interest rate models

Most recent term structure models include the possibility of regime-switching for short-term interest rates. Garcia and Perron [68] analysed the time series behaviour of U.S. real interest rates from 1961 to 1986 and found empirical evidence for jumps, which were caused by important structural events, such as a sudden rise in the oil price or an expected federal budget deficit. Hamilton [79] introduced changes in regimes by modelling parameters of an autoregression with a discrete Markov chain applied to a business cycle. His general time series model for changes in regimes (Hamilton [81]) follows a first order autoregression process, where the constant term and the autoregressive coefficient might be different for different regimes. In particular, the model for the process A_t takes the form

$$A_t = c_{N_t} + \Phi_{N_t} A_{t-1} + \varepsilon_t \,. \tag{5.4}$$

The error term ε_t is IID and follows an $N(0, \sigma^2)$ distribution. The different regimes are denoted by the subscript N_t, which is modelled as an outcome of an unobserved Markov chain with N states. This model is more flexible than deterministic autoregressive models since exogenous changes can be taken into account (e.g changes in the performance of the economy). Various regime-switching models for interest rates were inspired by Hamilton's model.

Gray [75] used a generalised regime-switching model for short-term interest rates, which allows mean-reversion and conditional heteroscedasticity. The model outperforms single-regime models in out-of-sample forecasting performance. Similar evidence for good performances of regime-switching short term interest rate models is described by Bansal and Zhou [10]. Here the efficient method of moments (Bansal et al. [9]) is used for the estimation of the model parameters. Amongst others, Naik and Lee [121] and Evans [64] include regime-switching in the volatility

term of short term interest rate models. Driffill, Kenc and Sola [45] found empirical evidence that regime shifts add more realism to interest rate models. They analyse the effects of regime-switching parameters in a Cox-Ingersoll-Ross term structure process and found that the possibility of regime shifts in the volatility parameter and the long-term mean of the process lead to a better fit of bond yields compared to models without regime-switching. A study by Smith [140] supports Markov switching models over stochastic volatility models, because the volatility seems to depend on the level of the short rate.

Landén [97] developed a hidden Markov model for short term interest rates, where drift and diffusion parameters are modulated by an underlying Markov process. In a model by Elliott, Hunter and Jamieson [53] the short rate process r is simply a function of a Markov chain and corresponding bond prices are observed in noise. A generalisation of Elliott et al.'s modelling framework, where the Markov chain's state could be time-dependent, is developed in Mamon [110]. Furthermore a closed-form solution for bond prices, where the underlying short rate is modelled by a mean reversion level governed by a continuous time Markov chain is derived in Elliott and Mamon [56], however the volatility process is constant in their model formulation.

An important problem in these models is the parameter estimation, which can be very complex since model parameters have to be estimated for each regime. Motivated by the empirical evidence supporting the possibility of regime shifts, an HMM for the short term interest rate is developed in this chapter, which generalises the model of Hull and White [91]. Whilst the paper of Elliott, Fischer and Platen [51] addresses the HMM filtering of a mean-reverting model, the derived filters are in continuous time setting and only a simulation is given. In this chapter, recursive estimates for the model parameters are provided using the approach discussed in chapter 4. We derive the exact adaptive filters for Markov chains observed in Gaussian noise, together with the jump process and occupation times. Employing further the EM algorithm, the recursive parameter estimates for the interest rate model are explicitly given. All calculations are made under an idealised measure $\bar{P}$, equivalently to the real world measure P.

From a practical point of view this recursive parameter estimation method is able to update parameters on-line, that is, new estimates of parameters are obtained as soon as a new set of data points becomes available. In contrast, many financial models are still heavily dependent on the static model fitting approach of regression method or maximum likelihood estimation (see Elliott, Hunter and Jamieson [53]).

5.2 Model description

In the proposed extended model for interest rates in this chapter consider the stochastic dynamics of r_t given in (5.2). Rearranging the formula leads to

$$dr_t = a_t[\delta_t - r_t]\,dt + u_t\,dW_t \tag{5.5}$$

with $\delta = \frac{\theta}{a}$. Equation (5.5) is a particular case of an Ornstein-Uhlenbeck process with mean reversion level δ. When the parameters are constant it has the solution

$$\begin{aligned} r_t &= r_0 e^{-at} + \int_0^t a\, e^{-a(t-u)}\delta\, du + \int_0^t e^{-a(t-u)} u\, dW_u \\ &= r_0 e^{-at} + (1 - e^{-at})\delta + u e^{-at} \int_0^t e^{au}\, dW_u \ . \end{aligned} \tag{5.6}$$

On the basis of the Hull-White specification we shall develop a hidden Markov model for the interest rate r. Suppose we observe the short rate r, where r is a data series in discrete time. Suppose further that the short rate process r_t can be proxied by the yield rates of T-bills with very short maturity observed in discrete time. An $N-$state discrete time Markov chain $\mathbf{x}_k$, which represents different states of the economy, is hidden in these observed values. As in the previous chapter, we work under the underlying probability space $(\Omega, \mathcal{F}, P)$, where $\mathbf{x}_k$ is a homogeneous Markov chain with finite state in discrete time ($k = 0, 1, ...$). The distribution of $\mathbf{x}_0$ is known and the state space of $\mathbf{x}_k$ is described by the canonical basis of $\mathbb{R}^N$, $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$. Let $\mathcal{F}_k^0 = \sigma\{\mathbf{x}_0, \ldots, \mathbf{x}_k\}$ be the $\sigma-$field generated by $\mathbf{x}_0, \ldots, \mathbf{x}_k$, and $\mathcal{F}_k^{\mathbf{x}}$ be the complete filtration generated by $\mathcal{F}_k^0$. Furthermore let $\mathcal{R}_k$ denote the complete filtration generated by r, so that $\mathcal{H}_k = \mathcal{F}_k^{\mathbf{x}} \vee \mathcal{R}_k$ is the global filtration generated by $\mathbf{x}$ and r.

The interest rate r with regime-switching parameters follows the stochastic process

$$dr_t = a(\mathbf{x}_t)[\delta(\mathbf{x}_t) - r_t]\,dt + u(\mathbf{x}_t)\,dW_t \tag{5.7}$$

for $r_0 \geq 0$ with $a(\mathbf{x}_t) = \langle \mathbf{a}, \mathbf{x}_t \rangle$, $\delta(\mathbf{x}_t) = \langle \boldsymbol{\delta}, \mathbf{x}_t \rangle$ and $u(\mathbf{x}_t) = \langle \boldsymbol{u}, \mathbf{x}_t \rangle$, where $\langle \cdot, \cdot \rangle$ is the usual Euclidean scalar product. All three parameters are governed by a Markov chain, which ensures, that the model is switching from one economic regime to another through time.

Consider the interest rate process over the time interval $[s, t]$. Then, if $t - s$ is small and $\mathbf{x}$ is constant over the interval, the solution of the stochastic process in equation (5.7), after invoking (5.6), is

$$\begin{aligned} r_t &= e^{-a(\mathbf{x}_s)(t-s)} r_s + \delta(\mathbf{x}_s)(1 - e^{-a(\mathbf{x}_s)(t-s)}) \\ &\quad + u(\mathbf{x}_s) e^{-a(\mathbf{x}_s)t} \int_s^t e^{a(\mathbf{x}_s)u} dW_u \,. \end{aligned} \tag{5.8}$$

The stochastic integral $e^{-a(\mathbf{x}_s)t} \int_s^t e^{a(\mathbf{x}_s)u} dW_u$ is normally distributed with mean zero and variance

$$\int_s^t e^{2a(\mathbf{x}_s)(u-t)} du = \frac{(1 - e^{-2a(\mathbf{x}_s)(t-s)})}{2a(\mathbf{x}_s)}.$$

From equation (5.8) the discrete-time representation of the interest rate process is derived as

$$r_{k+1} = \alpha(\mathbf{x}_k) r_k + \gamma(\mathbf{x}_k) + \xi(\mathbf{x}_k) w_{k+1} \tag{5.9}$$

where

$$\begin{aligned} \alpha(\mathbf{x}_k) &= e^{-a(\mathbf{x}_k)\Delta} \\ \gamma(\mathbf{x}_k) &= \delta(\mathbf{x}_k)(1 - e^{-a(\mathbf{x}_k)\Delta}) \end{aligned}$$

and

$$\xi(\mathbf{x}_k) = u(\mathbf{x}_k) \sqrt{\frac{1 - e^{-2a(\mathbf{x}_k)\Delta}}{2a(\mathbf{x}_k)}} \,.$$

Here, $\{\mathbf{x}_k\}$ is a discrete-time Markov chain and $\{w_k\}$ is a sequence of IID standard normal random variables. For this discrete-time version of the Hull-White model, optimal parameter estimates can be derived with the filtering techniques described in chapter 4.

Now let's assume that we have an observation process $\{y_k : k \in \mathbb{N}\}$ of the form

$$y_{k+1} = \alpha(\mathbf{x}_k)y_k + \gamma(\mathbf{x}_k) + \xi(\mathbf{x}_k)w_{k+1} \quad . \tag{5.10}$$

The filtrations generated by the processes are defined as $\mathcal{F}^y = \sigma(y_1, y_2, ...)$, $\mathcal{F}^{\mathbf{x}} = \sigma(\mathbf{x}_1, \mathbf{x}_2, ...)$ and $\mathcal{F} = \mathcal{F}^y \vee \mathcal{F}^{\mathbf{x}}$. The derivation of optimal parameter estimates shall be carried out under a reference probability measure under which the observation process and the Markov chain are IID. Following the technique described in section 2.3, the real-world measure is derived here from the reference probability measure for the discretised OU-process with regime-switching parameters. The Girsanov theorem in discrete time is utilised following chapter 8 of Elliott, Aggoun and Moore [49]. Let us define the measure P by $\frac{dP}{d\bar{P}} \mid_{\mathcal{F}_t} = \bar{\Lambda}_t$ with

$$\begin{aligned} \bar{\lambda}_l &= \exp\Bigg[-\frac{\langle \boldsymbol{\alpha}, \mathbf{x}_{l-1}\rangle y_{l-1} + \langle \boldsymbol{\gamma}, \mathbf{x}_{l-1}\rangle}{\langle \boldsymbol{\xi}, \mathbf{x}_{l-1}\rangle} \cdot \frac{y_l}{\langle \boldsymbol{\xi}, \mathbf{x}_{l-1}\rangle} \\ &\qquad - \frac{(\langle \boldsymbol{\alpha}, \mathbf{x}_{l-1}\rangle y_{l-1} + \langle \boldsymbol{\gamma}, \mathbf{x}_{l-1}\rangle)^2}{2\langle \boldsymbol{\xi}, \mathbf{x}_{l-1}\rangle^2}\Bigg] \end{aligned} \tag{5.11}$$

$$\bar{\Lambda}_l = \prod_{k=1}^{l} \bar{\lambda}_k \tag{5.12}$$

with $\bar{\Lambda}_0 = 1, \{\bar{\lambda}_l : l \in \mathbb{N}\}$ and $\{\bar{\Lambda}_l : l \in \mathbb{N}\}$. To show that the process $\{\bar{\Lambda}_l\}$ is an $\mathcal{F}$-martingale under $\bar{P}$, let us consider the conditional expectation

$$\bar{\mathrm{E}}[\bar{\Lambda}_{t+1} \mid \mathcal{F}_t] = \bar{\mathrm{E}}[\bar{\Lambda}_t\bar{\lambda}_{t+1} \mid \mathcal{F}_t] = \bar{\Lambda}_t\bar{\mathrm{E}}[\bar{\lambda}_{t+1} \mid \mathcal{F}_t].$$

Define the term in the exponent of $\bar{\lambda}_{t+1}$, as

$$I := -\frac{\alpha(\mathbf{x}_t)y_t + \gamma(\mathbf{x}_t)}{\xi(\mathbf{x}_t)} \cdot \frac{y_{t+1}}{\xi(\mathbf{x}_t)} - \frac{(\alpha(\mathbf{x}_t)y_t + \gamma(\mathbf{x}_t))^2}{2\xi(\mathbf{x}_t)^2}.$$

Then

$$\begin{aligned} \bar{\mathrm{E}}[I \mid \mathcal{F}_t] &= -\frac{1}{2}\Big(\frac{\alpha(\mathbf{x}_t)y_t + \gamma(\mathbf{x}_t)}{\xi(\mathbf{x}_t)}\Big)^2 \\ \overline{\mathrm{Var}}[I \mid \mathcal{F}_t] &= \Big(\frac{\alpha(\mathbf{x}_t)y_t + \gamma(\mathbf{x}_t)}{\xi(\mathbf{x}_t)}\Big)^2. \end{aligned}$$

Therefore $\bar{\mathrm{E}}\big[\exp(I) \mid \mathcal{F}_t\big] = \exp\big[\bar{E}\big[I \mid \mathcal{F}_t\big] + \frac{1}{2}\bar{\mathrm{Var}}\big[I \mid \mathcal{F}_t\big]\big] = 1$ and $\bar{\mathrm{E}}\big[\bar{\Lambda}_{t+1} \mid \mathcal{F}_t\big] = \bar{\Lambda}_t$. The process $\{\bar{\Lambda}_l\}$ is an $\mathcal{F}$-martingale under $\bar{P}$.

Furthermore $\bar{\Lambda}\Lambda = 1$ with

$$\Lambda_l = \prod_{k=1}^{l} \exp\left[\frac{\langle \boldsymbol{\alpha}, \mathbf{x}_{l-1}\rangle y_{l-1} + \langle \boldsymbol{\gamma}, \mathbf{x}_{l-1}\rangle}{\langle \boldsymbol{\xi}, \mathbf{x}_{l-1}\rangle^2} \cdot y_l + \frac{(\langle \boldsymbol{\alpha}, \mathbf{x}_{l-1}\rangle y_{l-1} + \langle \boldsymbol{\gamma}, \mathbf{x}_{l-1}\rangle)^2}{2\langle \boldsymbol{\xi}, \mathbf{x}_{l-1}\rangle^2}\right].$$

The following section describes the optimal parameter estimation for the observation process. Invoking the Bayes' theorem, the recursive parameter estimates expressed under the real-world measure P are calculated.

5.3 Parameter estimation

The set of parameters ρ, which determines the regime-switching interest rate model is

$$\rho = \{\pi_{ji}, \alpha_i, \gamma_i, \xi_i, 1 \leq i, j \leq N\}. \tag{5.13}$$

The EM algorithm (see subsection 3.3.1) is applied to determine the optimal estimate for each parameter in the set ρ. Initial values for the EM algorithm are assumed to be given. Starting from these values updated parameter estimates are derived which maximise the conditional expectation of the log-likelihoods. The recursive filters from processes of the Markov chain are employed, which were derived in chapter 4. Filters for the jump process, the occupation time process as well as for the auxiliary processes are required for the optimal parameter estimates. Write $\hat{H}_k = E[H_k|\mathcal{F}_k^y]$ for any adapted process H.

The EM algorithm involves a change of measure from P^ρ to $P^{\widehat{\rho}}$. Under P^ρ, $\mathbf{x}$ is a Markov chain with transition matrix $\boldsymbol{\Pi} = (\pi_{ji})$. Under $P^{\widehat{\rho}}$, $\mathbf{x}$ is still a Markov chain with transition matrix $\widehat{\boldsymbol{\Pi}} = (\widehat{\pi}_{ji})$. Thus, $P^{\widehat{\rho}}(\mathbf{x}_{k+1} = \mathbf{e}_j|\mathbf{x}_k = \mathbf{e}_i) = \widehat{\pi}_{ji}$. Therefore, $\widehat{\pi}_{ji} \geq 0$ and $\sum_{j=1}^n \widehat{\pi}_{ji} = 1$. To find an estimate for the transition probability matrix $\boldsymbol{\Pi} = (\pi_{ji})$, where $\sum_{i=1}^n \pi_{ji} = 1$ we consider the Radon-Nikodým derivative

$$\left.\frac{d\hat{P}}{dP}\right|_{\mathcal{F}_k^y} := \Lambda_k^\pi = \prod_{l=1}^{k}\left(\prod_{c,d=1}^{n}\left(\frac{\hat{\pi}_{dc}}{\pi_{dc}}\right)^{\langle \mathbf{x}_l, \mathbf{e}_d\rangle\langle \mathbf{x}_{l-1}, \mathbf{e}_c\rangle}\right)$$
$$\text{with} \quad \Lambda_0 = 1 \quad \text{and} \quad \mathcal{F}_k^y = \sigma(y_1, y_2, \ldots, y_k).$$

With $\hat{J}$, $\hat{O}$ and $\hat{T}$ denoting the best estimates for the processes J, O and T, respectively, the optimal parameter estimates $\hat{\pi}_{ji}, \hat{\alpha}_i, \hat{\gamma}_i, \hat{\xi}_i$ are given by

$$\hat{\pi}_{ji} = \frac{\hat{J}_k^{ji}}{\hat{O}_k^i} = \frac{\eta_k(J_k^{ji})}{\eta_k(O_k^i)} \tag{5.14}$$

$$\hat{\alpha}_i = \frac{\hat{T}_k^i(y_{k+1}, y_k) - \hat{T}_k^i(y)\gamma_i}{\hat{T}_k^i(y^2)} = \frac{\eta_k(T_k^i(y_{k+1}, y_k)) - \eta_k(T_k^i(y))\gamma_i}{\eta_k(T_k^i(y^2))} \tag{5.15}$$

$$\hat{\gamma}_i = \frac{\hat{T}_{k+1}^i(y) - \hat{T}_k^i(y)\hat{\alpha}_i}{\hat{O}_k^i} = \frac{\eta_{k+1}(T_{k+1}^i(y)) - \eta_k(T_k^i(y))\hat{\alpha}_i}{\eta_k(O_k^i)} \tag{5.16}$$

$$\begin{aligned}\hat{\xi}_i &= \frac{\hat{T}_{k+1}^i(y^2) + \hat{\alpha}_i^2\hat{T}_k^i(y^2) + \hat{\gamma}_i^2\hat{O}_k^i - 2\hat{\alpha}_i\hat{T}_k^i(y_{k+1}, y_k)}{\hat{O}_k^i} \\ &\quad - \frac{2\hat{\gamma}_i\hat{T}_{k+1}^i(y) + 2\hat{\alpha}_i\hat{\gamma}_i\hat{T}_k^i(y)}{\hat{O}_k^i} \\ &= \frac{\eta_{k+1}(T_{k+1}^i(y^2)) + \hat{\alpha}_i^2\eta_k(T_k^i(y^2)) + \hat{\gamma}_i^2\eta_k(O_k^i) - 2\hat{\alpha}_i\eta_k(T_k^i(y_{k+1}, y_k))}{\eta_k(O_k^i)} \\ &\quad - \frac{2\hat{\gamma}_i\eta_{k+1}(T_{k+1}^i(y)) + 2\hat{\alpha}_i\hat{\gamma}_i\eta_k(T_k^i(y))}{\eta_k(O_k^i)} \quad .\end{aligned} \tag{5.17}$$

Note here, that all the recursive parameter estimates can be expressed fully through the vector recursions derived in the previous chapter, equations (4.7), (4.9) and (4.11) with the aid of relation (4.3). The proofs of equations (5.15) – (5.17) are provided in Appendices C.1 - C.3. As in Elliott, Sick and Stein [57] the first parameter α is updated using the parameter estimate γ from the previous calculated optimal parameter set. However, once α is updated, this new optimal parameter estimate is used for updating the remaining parameters. Since the optimal parameter estimates are obtained through recursive filters, estimates can be updated whenever new information is available in the data set. We therefore have a self-calibrating model, which is an on-line estimation scheme for parameters in a discretised regime-switching OU-process setting.

5.4 Implementation

The filtering method for a regime-switching interest rate model is implemented on a data set consisting of 30-day Treasury-bill (T-bill) rates. T-bills are instruments with a maturity of one year or less which do not pay interest prior to maturity.

They are regarded as one of the most risk-free investments. Short term Treasury-bill rates are therefore useful as a proxy for the short term interest rates.

The data set used for the implementation is compiled by the Bank of Canada and consists of 2880 data points, which are daily 30-day T-bill yields between the 3rd of January 1996 and the 15th of June 2007. Our optimal parameter estimation method processes the data in batches of 20 data points. The parameters are roughly updated monthly, for this data set, the parameters are updated 144 times. The parameter estimation technique offers a choice of frequency for parameter updates, other frequencies for data sets can be easily adopted. This gives more flexibility to users in updating the model parameters either on a more frequent or on a less frequent basis depending on the dictates of the financial market.

A preliminary analysis of the actual data reveals that the evolution of the T-bill rates undergoes several distinct regimes characterised by states with high and low means as well as high and low standard deviations. The regime-switching model is proposed to capture this particular behaviour. In Figures 5.1 and 5.2 together with Table 5.1 possible segregations of actual data into either two or three states based on combinations between high or low means and standard deviations are displayed. The Canadian T-bill rate is relatively stable over monthly intervals, therefore a monthly update of parameters is deemed sufficient in this data set. The largest decline in T-bill rates can be seen in 2001. The events of September 11th, 2001 lead to a crisis in economies worldwide, and is clearly visible in the short term T-bill rates. The Bank of Canada lowered its key overnight rate on the 17th of September 2001 by 0.5% and continued to lower interest rates into early 2002. The aim of lowering interest rates was to restore consumer and investor confidence and since the T-bill rates were already declining in 2001 prior to September 11th, this lead to an overall decline from 5.44% in January 2001 to 1.38% in January 2002. The Figures below show possible state segregations. As soon as the algorithm is run, it picks up information from the historical data and learns to adjust quickly to the trends of the data series. A prior state segregation is not necessary, however, the number of states has to be chosen. A discussion on choosing the number of states is presented in subsection 5.4.3.

Figure 5.1: Possible segregation of historical T-bill rates into 2 states

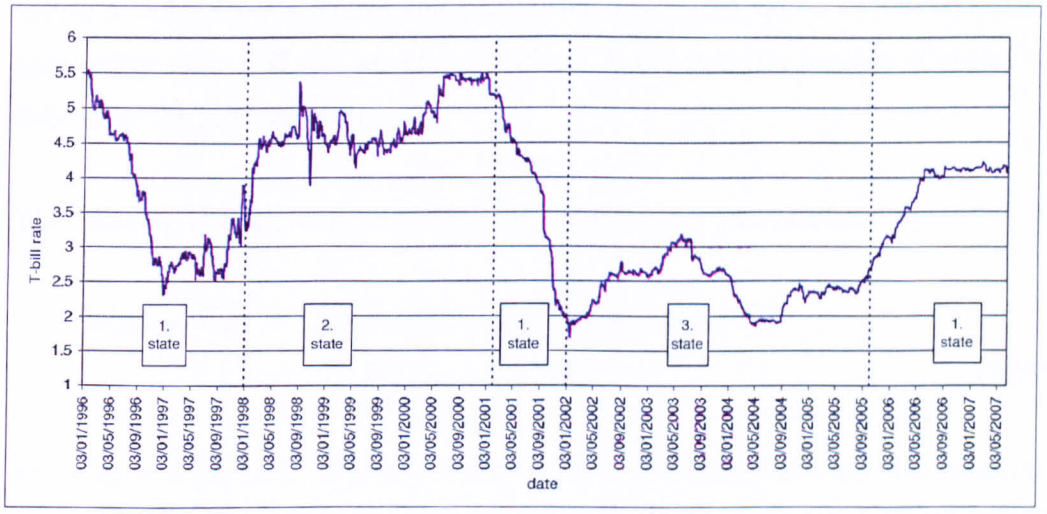


Figure 5.2: Possible segregation of historical T-bill rates into 3 states

Since the parameter estimates are calculated through the EM algorithm, initial values for the implementation have to be chosen. These values must be reasonable so that the algorithm yields local maxima. In order to be systematic, the initial values for the algorithm are found by employing a least-square method on the first few data points. In particular, the discretised interest rate function specified in equation (5.9) under a 1-state setting is fitted to the first 50 data points of the

1st state	2nd state
Jan 1996 - Sept 1996 mean: 4.63 std: 0.50	Oct 1996 - Dec 1997 mean: 2.89 std: 0.30
Jan 1998 - Sept 2001 mean: 4.68 std: 0.47	Oct 2001 - Jan 2006 mean: 2.46 std: 0.36
Feb 2006 - June 2007 mean: 4.00 std: 0.23	

1st state	2nd state	3rd state
Jan 1996 - Dec 1997 mean: 3.55 std: 0.93		
Feb 2001 - Dec 2001 mean: 3.81 std: 0.99	Jan 1998 - Jan 2001 mean: 4.76 std: 0.43	Jan 2002 - Sept 2005 mean: 2.43 std: 0.33
Oct 2005 - June 2007 mean: 3.83 std: 0.44		

Table 5.1: Segregation of the period of actual data into 2 or 3 states

sample. The resulting parameter estimates are used as rough guides for the initial values of the parameters α, γ and ξ. The least square parameter estimation was carried out using the MATLAB function 'lsqcurvefit' and this gives the parameter values $\alpha = 0.9966, \gamma = 0.0038$ and $\xi = 0.0105$. The initial values for the transition probability matrix $\mathbf{\Pi}$ are set to $1/N$, where N denotes the number of states as defined in the implementation.

The one-step ahead predicted yields of the T-bill rates are calculated by

$$\begin{aligned} \mathrm{E}\big[y_{k+1} \mid \mathcal{F}_k\big] &= \mathrm{E}\big[\alpha(\mathbf{x}_k)y_k + \gamma(\mathbf{x}_k) + \xi(\mathbf{x}_k)w_{k+1} \mid \mathcal{F}_k\big] \\ &= \langle \boldsymbol{\alpha}, \mathbf{\Pi}\hat{\mathbf{x}}_k\rangle y_k + \langle \boldsymbol{\gamma}, \mathbf{\Pi}\hat{\mathbf{x}}_k\rangle \end{aligned} \tag{5.18}$$

where $\hat{\mathbf{x}}_k = E\big[\mathbf{x}_k \mid \mathcal{F}_k\big]$. Figure 5.3 shows the actual time series and the resulting one-step ahead forecasts generated by a 2-state HMM-based interest rate model between 1996 and 2007.

As a preliminary, the number of states for the HMM is imposed on the implementation considering the realistic features of the actual data seen in the preliminary data analysis. Following the data segregation displayed in Table 5.1 we generate 1-step ahead forecasts using the 1-, 2- and 3-state HMM-based models. For the 2-state case, it is evident that the forecasts and actual data in Figure 5.3 are very close to each other. Forecasts using a 4-state HMM-based interest rate model were generated to see if there is any further improvement that can still be gained. How-

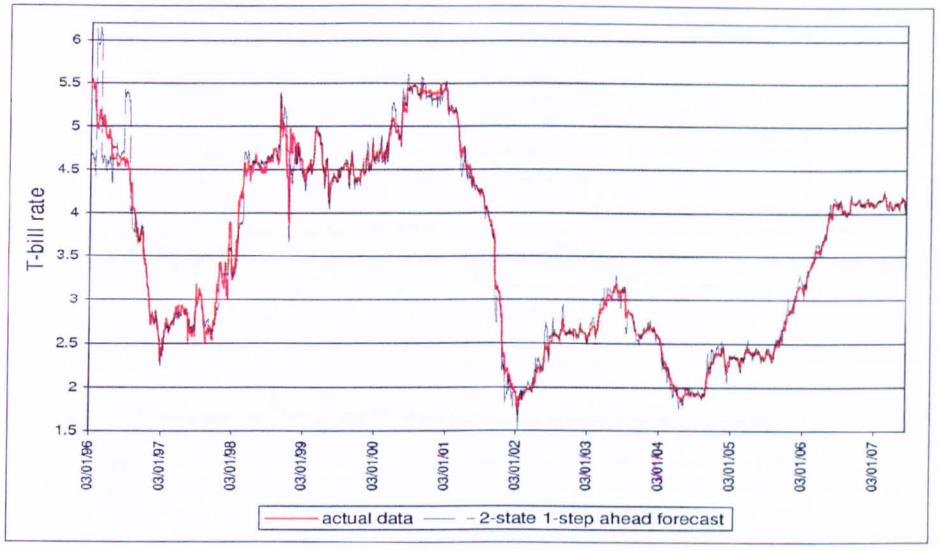


Figure 5.3: Plot of actual and one-step ahead forecasts generated by a 2-state HMM-based interest rate model

ever, we did not find evidence of this. The evolution of the parameters after 144 passes under the 2-state and 3-state HMM-based interest rate model is shown in Figures 5.4 and 5.5. The MATLAB code for the implementation of the 3-state HMM for interest rates is stated in Appendix C.4.

5.4.1 Fisher information

The sensitivity of the model with respect to changes in the parameters can be quantified through the Fisher information $\mathcal{I}$. We compute the Fisher information $\mathcal{I}$ for each parameter α, γ, ξ and π_{ij} after each parameter update. The Fisher information is defined as the variance of the score and can be calculated through

$$\mathcal{I}(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right]. \tag{5.19}$$

In the following the Fisher information is derived for the estimated parameters.

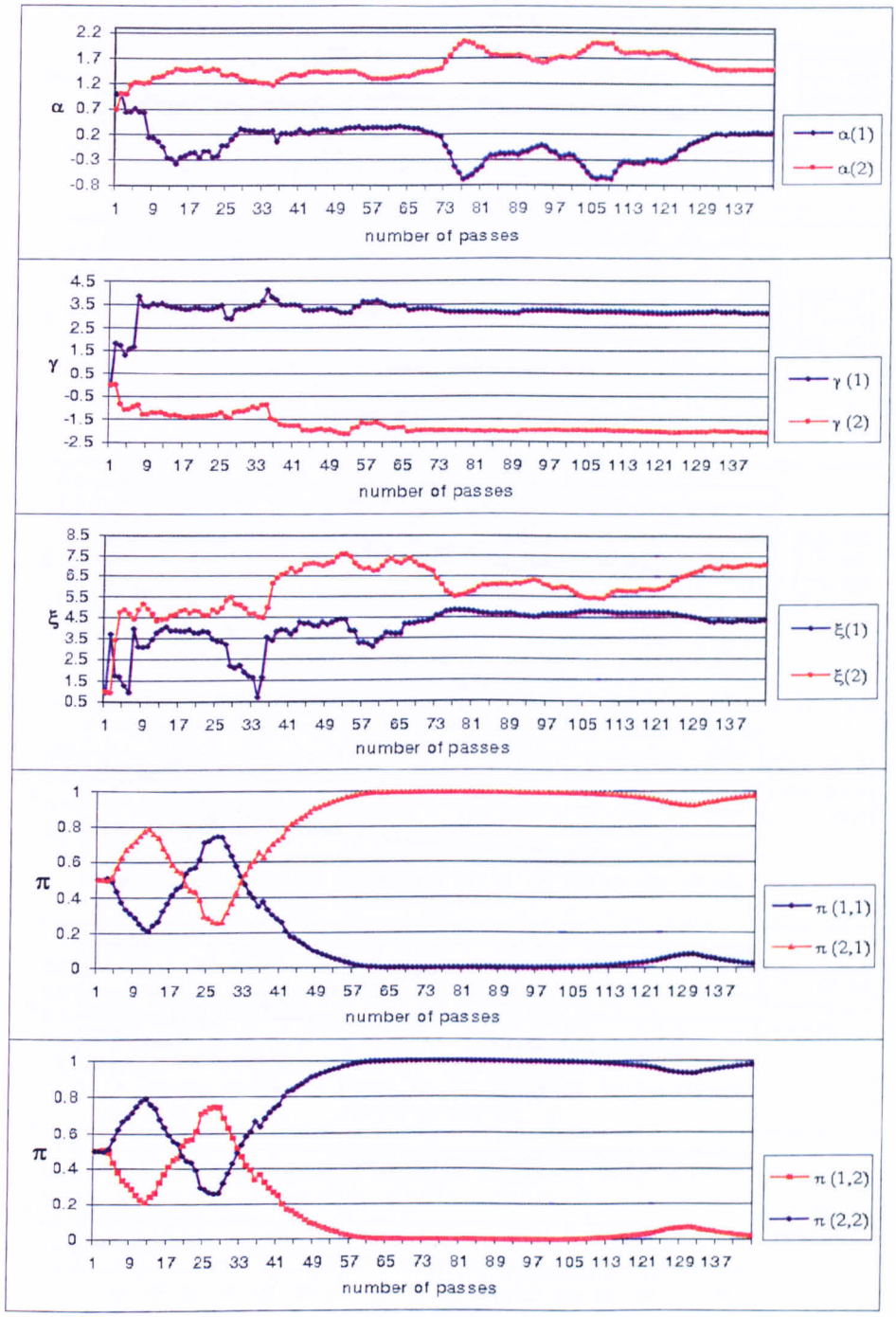


Figure 5.4: Evolution of estimates for the parameters $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\xi}$ and the transition probabilities for a 2-state HMM-based interest rate model

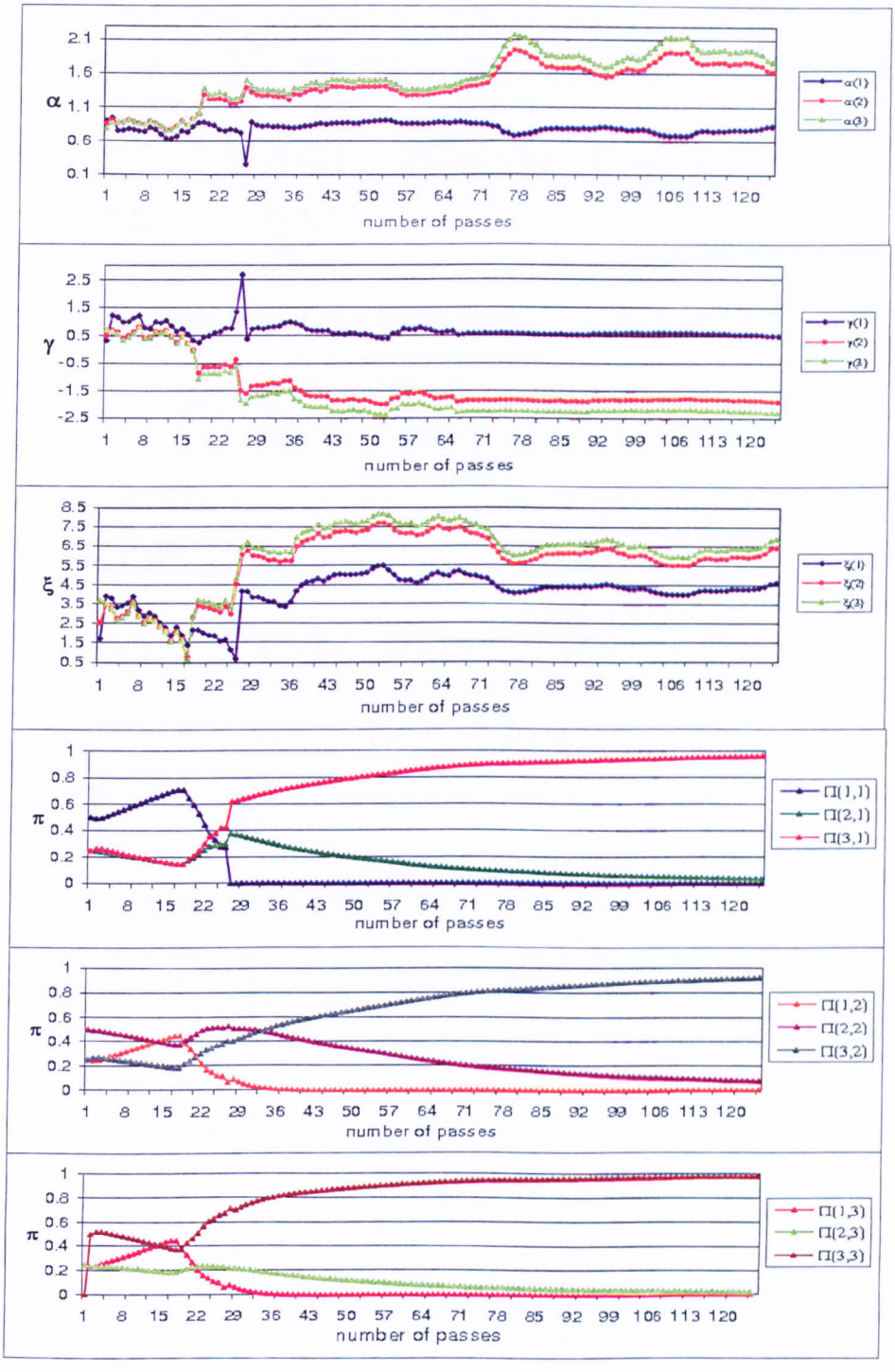


Figure 5.5: Evolution of estimates for the parameters $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\xi}$ and the transition probabilities for a 3-state HMM-based interest rate model

Fisher information for α :

$$\begin{aligned}
\log \Lambda_k^{\alpha} &= \sum_{i=1}^{n}\left(-T_k^i(y^2)\hat{\alpha}_i^2 + 2T_k^i(y_{k+1},y_k)\hat{\alpha}_i - 2T_k^i(y)\gamma_i\hat{\alpha}_i\right)/2\xi_i^2 + R(\alpha_i) \\
\frac{\partial \log \Lambda_k^{\alpha}}{\partial \hat{\alpha}_i} &= \left(-2\hat{\alpha}_i T_k^i(y^2) + 2T_k^i(y_{k+1},y_k) - 2T_k^i(y)\gamma_i\right)/2\xi_i^2 \\
\frac{\partial^2 \log \Lambda_k^{\alpha}}{\partial \hat{\alpha}_i^2} &= -\frac{2T_k^i(y^2)}{2\xi_i^2}.
\end{aligned}$$

Therefore the Fisher information for parameter α_i is given by

$$\mathcal{I}_k(\alpha_i) = -E\left[-\frac{2T_k^i(y^2)}{2\xi_i^2}\right] = \frac{2\hat{T}_k^i(y^2)}{2\xi_i^2}. \tag{5.20}$$

Fisher information for γ :

$$\begin{aligned}
\log(\Lambda_k^{\gamma}) &= \sum_{i=1}^{n}\left(-O_k^i\hat{\gamma}_i^2 + 2T_{k+1}^i(y)\hat{\gamma}_i - 2T_k^i(y)\alpha_i\hat{\gamma}_i\right)/2\xi_i^2 + R(\gamma_i) \\
\frac{\partial \log \Lambda_k^{\gamma}}{\partial \hat{\gamma}_i} &= \left(-O_k^i\hat{\gamma}_i + 2T_{k+1}^i(y) - 2T_k^i(y)\alpha_i\right)/2\xi_i^2 \\
\frac{\partial^2 \log \Lambda_k^{\gamma}}{\partial \hat{\gamma}_i^2} &= \frac{-O_k^i}{2\xi_i^2}.
\end{aligned}$$

The Fisher information of parameter γ_i is then given by

$$\mathcal{I}_k(\gamma_i) = -E\left[-\frac{O_k^i}{2\xi_i^2}\right] = \frac{\hat{O}_k^i}{2\xi_i^2}. \tag{5.21}$$

Fisher information for ξ :

$$\begin{aligned}
\log \Lambda_k^{\xi} &= \sum_{i=1}^{n}\Bigg(-\log\hat{\xi}_i O_k^i - \frac{T_{k+1}^i(y^2) + \alpha_i^2 T_k^i(y^2) + \gamma_i^2 O_k^i}{2\hat{\xi}_i^2} \\
&\qquad + \frac{\alpha_i T_k^i(y_{k+1},y_k) + \gamma_i T_{k+1}^i(y) - \alpha_i\gamma_i T_k^i(y)}{\hat{\xi}_i^2}\Bigg) + R(\xi) \\
\frac{\partial \log \Lambda_k^{\xi}}{\partial \hat{\xi}_i} &= -\frac{O_k^i}{\hat{\xi}_i} + \hat{\xi}_i^{-3}\left[T_{k+1}^i(y^2) + \alpha_i^2 T_k^i(y^2) + O_k^i\gamma_i^2\right] \\
&\quad -2\hat{\xi}_i^{-3}\left[T_k^i(y_{k+1},y_k)\alpha_i + T_{k+1}^i(y)\gamma_i - T_k^i(y)\alpha_i\gamma_i\right] \\
\frac{\partial^2 \log \Lambda_k^{\xi}}{\partial \hat{\xi}_i^2} &= O_k^i\hat{\xi}_i^{-2} - 3\hat{\xi}_i^{-4}\left[T_{k+1}^i(y^2) + \alpha_i^2 T_k^i(y^2) + O_k^i\gamma_i^2 - 2T_k^i(y_{k+1},y_k)\alpha_i\right. \\
&\quad \left. -2T_{k+1}^i(y)\gamma_i + 2T_k^i(y)\alpha_i\gamma_i\right].
\end{aligned}$$

The Fisher information of parameter ξ_i is then given by

$$\begin{aligned}\mathcal{I}_k(\xi_i) &= -\hat{O}_k^i\hat{\xi}_i^{-2} + 3\hat{\xi}_i^{-4}\big[\hat{T}_{k+1}^i(y^2) + \alpha_i^2\hat{T}_k^i(y^2) + \hat{O}_k^i\gamma_i^2 - 2\hat{T}_k^i(y_{k+1}, y_k)\alpha_i \\ &\quad -2\hat{T}_{k+1}^i(y)\gamma_i + 2\hat{T}_k^i(y)\alpha_i\gamma_i\big]. \end{aligned} \tag{5.22}$$

Fisher information for π in the 2-state case:

$$\begin{aligned}\log\Lambda_k^\pi &= \sum_{i,j=1}^{2}\big(J_k^{ji}\log\hat{\pi}_{ji}(k) + R(\pi_{ji})\big) \\ &= J_k^{11}\log\hat{\pi}_{11}(k) + J_k^{12}\log\hat{\pi}_{12}(k) + J_k^{21}\log(1-\hat{\pi}_{11}(k)) \\ &\quad + J_k^{22}\log(1-\hat{\pi}_{12}(k)) + \sum_{i,j=1}^{2} R(\pi_{ji})\end{aligned}$$

We note that $\hat{\pi}_{21} = 1 - \hat{\pi}_{11}$ and $\hat{\pi}_{22} = 1 - \hat{\pi}_{12}$. The Fisher information is calculated from the partial derivatives of the log-likelihood with respect to the parameters $\hat{\pi}_{11}, \hat{\pi}_{12}, 1-\hat{\pi}_{11}$ and $1-\hat{\pi}_{12}$. The partial derivatives with respect to $\hat{\pi}_{11}$ are

$$\begin{aligned}\frac{\partial\log\Lambda_k^\pi}{\partial\hat{\pi}_{11}} &= \frac{J_k^{11}}{\hat{\pi}_{11}(k)} - \frac{J_k^{21}}{1-\hat{\pi}_{11}(k)} \\ \frac{\partial^2\log\Lambda_k^\pi}{\partial\hat{\pi}_{11}^2} &= -\frac{J_k^{11}}{(\hat{\pi}_{11}(k))^2} - \frac{J_k^{21}}{(1-\hat{\pi}_{11}(k))^2}.\end{aligned}$$

The Fisher information of parameter π_{11} is then

$$\mathcal{I}_k(\pi_{11}) = \frac{\hat{J}_k^{11}}{(\hat{\pi}_{11}(k))^2} + \frac{\hat{J}_k^{21}}{(1-\hat{\pi}_{11}(k))^2}. \tag{5.23}$$

The partial derivatives of the log-likelihood for π_{12} are

$$\begin{aligned}\frac{\partial\log\Lambda_k^\pi}{\partial\hat{\pi}_{12}} &= \frac{J_k^{12}}{\hat{\pi}_{12}(k)} - \frac{J_k^{22}}{1-\hat{\pi}_{12}(k)} \\ \frac{\partial^2\log\Lambda_k^\pi}{\partial\hat{\pi}_{12}^2} &= -\frac{J_k^{12}}{(\hat{\pi}_{12}(k))^2} - \frac{J_k^{22}}{(1-\hat{\pi}_{12}(k))^2}\end{aligned}$$

and the Fisher information of parameter π_{12} is then given by

$$\mathcal{I}_k(\pi_{12}) = \frac{\hat{J}_k^{12}}{(\hat{\pi}_{12}(k))^2} + \frac{\hat{J}_k^{22}}{(1-\hat{\pi}_{12}(k))^2}. \tag{5.24}$$

For the parameter π_{21}, the Fisher information is calculated from the partial derivatives with respect to $1-\hat{\pi}_{11}$.

$$\begin{aligned}\frac{\partial \log \Lambda_k^{\pi}}{\partial(1-\hat{\pi}_{11})} &= \frac{J_k^{21}}{1-\hat{\pi}_{11}(k)} - \frac{J_k^{11}}{\hat{\pi}_{11}(k)} \\ \frac{\partial^2 \log \Lambda_k^{\pi}}{\partial(1-\hat{\pi}_{11})^2} &= -\frac{J_k^{21}}{(1-\hat{\pi}_{11}(k))^2} - \frac{J_k^{11}}{\hat{\pi}_{11}(k)^2}\end{aligned}$$

and is then given by

$$\mathcal{I}_k((1-\hat{\pi}_{11})) = \frac{\hat{J}_k^{11}}{\hat{\pi}_{11}(k)^2} + \frac{\hat{J}_k^{21}}{(1-\hat{\pi}_{11}(k))^2}. \tag{5.25}$$

The Fisher information of parameter π_{22} is calculated through

$$\begin{aligned}\frac{\partial \log \Lambda_k^{\pi}}{\partial(1-\hat{\pi}_{12})} &= \frac{J_k^{22}}{1-\hat{\pi}_{12}(k)} - \frac{J_k^{12}}{\hat{\pi}_{12}(k)} \\ \frac{\partial^2 \log \Lambda_k^{\pi}}{\partial(1-\hat{\pi}_{12})^2} &= -\frac{J_k^{22}}{(1-\hat{\pi}_{12}(k))^2} - \frac{J_k^{12}}{\hat{\pi}_{12}(k)^2}\end{aligned}$$

which leads to

$$\mathcal{I}_k((1-\hat{\pi}_{12})) = \frac{J_k^{12}}{\hat{\pi}_{12}(k)^2} + \frac{\hat{J}_k^{22}}{(1-\hat{\pi}_{12}(k))^2}. \tag{5.26}$$

Tables 5.4.1 exhibits the square root of the diagonal elements of the Fisher information matrix for the last 45 algorithm passes, namely passes 100 to 144. This quantity is a measure for the sensitivity of the model with respect to small changes in the corresponding parameters. If the sensitivity of a parameter is large then one can expect a small asymptotic variance for that parameter. In this case, the square root of the diagonal elements of the Fisher information matrix is larger for the parameter values in the second state. Therefore the variance of parameter estimates in the first state is expected to be higher. The transition probabilities in this particular data set indicate a low probability for being in state 1, the parameter estimates for that state are therefore less stable than those for the state 2. It has to be emphasised here, that the Fisher information is simply calculated separately for each parameter, cross-variations are not taken into account. This is due to the fact that the optimal parameter estimates in the model are derived from separate maximum likelihood functions (see appendices C.1-C.3) for each parameter and are

algorithm pass	α(1)	α(2)	γ(1)	γ(2)	ξ(1)	ξ(2)	π(1,1)	π(1,2)	π(2,1)	π(2,2)
100	3.6519	42.289	0.9837	11.392	4.3711	24.247	98.290	1477.9	96.290	1477.9
101	3.6526	42.241	1.0680	12.358	4.4399	24.748	102.93	1569.1	102.93	1569.1
102	3.5905	41.785	1.1356	13.217	4.4266	24.635	107.56	1601.5	107.56	1601.5
103	3.5477	41.785	1.2163	14.328	4.4395	24.674	113.57	1653.3	113.57	1653.3
104	3.5097	41.370	1.2803	15.099	4.4757	24.053	117.24	1655.5	117.24	1655.5
105	3.5504	40.647	1.3193	15.106	4.5874	23.577	116.16	1561.1	116.16	1561.1
106	3.7108	40.108	1.3530	14.624	4.8037	23.110	112.50	1419.8	112.50	1419.8
107	3.9476	40.456	1.4496	14.856	5.0926	23.571	114.40	1378.0	114.40	1378.0
108	4.1208	40.502	1.5291	15.030	5.3345	23.525	115.25	1324.0	115.25	1324.0
109	4.4885	39.605	1.5158	13.396	5.6783	22.632	104.52	1074.9	104.52	1074.9
110	5.1382	40.176	1.5941	12.466	6.3534	22.390	100.14	938.23	100.14	938.23
111	5.7735	40.728	1.7149	12.098	7.1044	22.986	99.801	875.87	99.801	875.87
112	6.2238	41.109	1.8489	12.214	7.6102	23.530	101.53	853.08	101.53	853.08
113	6.5158	41.284	1.9781	12.535	7.9305	23.864	103.81	838.70	103.81	838.70
114	6.7945	41.049	2.0580	12.434	8.3470	23.423	102.85	781.84	102.85	781.84
115	7.1833	41.127	2.1897	12.539	8.7436	23.559	104.06	753.87	104.06	753.87
116	7.6989	40.808	2.2422	11.886	9.2760	22.661	99.852	666.44	99.852	666.44
117	8.2076	41.198	2.4249	12.172	9.9970	23.311	103.14	665.18	103.14	665.18
118	8.5958	40.877	2.5378	12.069	10.393	23.144	102.61	628.27	102.61	628.27
119	8.9364	41.011	2.7002	12.392	10.804	23.232	105.09	619.54	105.09	619.54
120	9.3367	40.587	2.7825	12.096	11.319	22.680	103.14	571.51	103.14	571.51
121	10.077	40.545	2.8960	11.654	12.126	22.569	100.94	521.81	100.94	521.81
122	10.868	40.443	3.0130	11.213	12.842	22.088	98.799	478.53	98.799	478.53
123	12.319	40.741	3.0803	10.189	14.476	21.577	93.722	409.45	93.722	409.45
124	13.533	41.339	3.2710	9.9944	15.684	22.392	94.062	397.08	94.062	397.08
125	14.828	41.663	3.3231	9.3373	17.310	22.124	90.446	355.47	90.446	355.47
126	15.795	41.853	3.4190	9.0612	18.268	22.476	88.893	341.40	88.893	341.40
127	16.673	42.110	3.4415	8.8920	19.361	22.357	88.490	321.14	88.490	321.14
128	17.403	42.474	3.4542	8.4308	20.171	22.501	84.637	311.00	84.637	311.00
129	17.919	42.681	3.4244	8.1568	20.898	22.598	82.454	301.47	82.454	301.47
130	18.789	43.419	3.3363	7.7100	21.910	22.330	78.921	285.65	78.921	285.65
131	18.564	43.812	3.1933	7.5364	22.219	22.971	76.125	292.15	76.125	292.15
132	17.655	43.973	3.0616	7.6254	21.065	23.726	74.822	311.48	74.822	311.48
133	16.731	44.087	2.9629	7.8076	19.067	23.672	75.279	330.40	75.279	330.40
134	16.575	44.616	2.8340	7.6267	19.446	23.075	74.964	327.73	74.964	327.73
135	15.744	44.491	2.6929	7.8100	18.475	23.483	73.781	347.82	73.781	347.82
136	14.927	44.527	2.5698	7.8656	17.680	23.812	73.652	367.77	73.652	367.77
137	14.310	44.803	2.4570	7.8929	16.985	23.514	73.314	383.61	73.314	383.61
138	13.666	44.954	2.3521	7.7373	15.559	23.268	74.113	404.04	74.113	404.04
139	13.084	44.789	2.2320	7.6408	15.215	23.207	73.739	418.72	73.739	418.72
140	12.531	44.936	2.1085	7.5809	14.598	23.327	72.501	436.46	72.501	436.46
141	11.774	45.044	2.0088	7.6844	13.649	23.751	72.812	471.42	72.812	471.42
142	11.133	45.020	1.9151	7.7444	12.775	23.821	73.205	496.93	73.205	496.93
143	10.660	45.161	1.8231	7.7235	12.106	23.318	73.267	516.87	73.267	516.87
144	10.161	45.237	1.7355	7.7264	11.622	23.299	73.552	545.18	73.552	545.18

Table 5.2: Sensitivity analysis for changes in parameters in a 2-state HMM-based interest rate model

	MdAPE	MdRAE	MSE
1-state MC	0.0145	1.1306	0.0371
2-state MC	0.0126	0.9876	0.0269
3-state MC	0.0128	0.9928	0.0513

Table 5.3: Results of error analysis for the 1-, 2- and 3-state HMM-based interest rate model

updated one after the other. The Fisher information matrix is therefore not known as a whole, the values reported in table 5.4.1 are the square roots of the diagonal elements. One possible way to analyse the goodness of estimates further is to test the optimal parameter estimates in separate runs of the EM algorithm. The optimal parameter estimates found here can be chosen as initial values for running the EM algorithm without filtering on this data set. The EM algorithm leads to local maxima and should therefore give similar parameter estimates as previously found. This method shall be explored further in future work.

5.4.2 Assessment of Predicted Yields

The goodness of fit of the one-step ahead forecasts is assessed using the error measures defined in subsection 3.6.2. The median absolute percentage error (MdAPE) and the median relative absolute error (MdRAE) for the 1-, 2- and 3-state HMM-based models are evaluated and the mean square error (MSE) is also calculated. The forecasts of the models are then compared using these 3 criteria. The results of this error analysis are presented in Table 5.3.

The MdRAE, MdAPE and MSE values clearly indicate that the 2-state HMM-based interest rate model has the best fit in comparison with the 1- and 3-state HMM-based models. We note though that the error differences between the 2-state and 3-state models are not that significant.

5.4.3 The number of regimes

Selecting the number of states for the HMM-based interest rate model can be done with a penalised likelihood criteria, which is a standard procedure when choosing between nested models (see [133] and [85]). The optimal number of regimes for the proposed HMM-based interest rate model is determined by applying the Akaike information criterion (AIC) [1]. The derivation of this model selection criterion is based on the Kullback-Leibler information and utilises the log-likelihood function of the model together with a penalty term, which is the number of model parameters. Specifically, the AIC is given by

$$AIC = -2\log(\mathcal{L}(\rho)) + 2s,$$

where $\mathcal{L}(\rho)$ is the likelihood function of the model and s denotes the number of parameters. The model which minimises the AIC is preferred over the others. The log-likelihood function of the parameter set ρ (see (5.13)) for the observation process y in each pass is given by

$$\begin{aligned}\mathcal{L}(\rho) &= \sum_{k=1}^{K} \log f_{Y_k|Y_{k-1}}(y_k \mid y_{k-1};\rho) \\ &= \sum_{k=1}^{K}\left(-\frac{1}{2}\log(2\Pi\xi(\mathbf{x}_k)^2) - \frac{(y_k - \alpha(\mathbf{x}_k)y_{k-1} - \gamma(\mathbf{x}_k))^2}{2\xi(\mathbf{x}_k)^2}\right)\end{aligned}$$

where $f_{Y_k|Y_{k-1}}$ denotes the density of observation k conditional on the preceding observation $k-1$ and K is the number of observations in each pass.

Figure 5.6 shows the calculated AIC for 1-, 2-, 3- and 4-state HMM-based models after each pass. The parameter values used to obtain the forecast in the current pass generate the log-likelihood function and the AIC in that given pass. Due to this information criterion the best performing model that captures the dynamics of this interest rate series is the 2-state model. This finding is consistent with the results of the error analysis given in Table 5.3. Therefore a 2-state model is sufficient to model these T-bill rates.

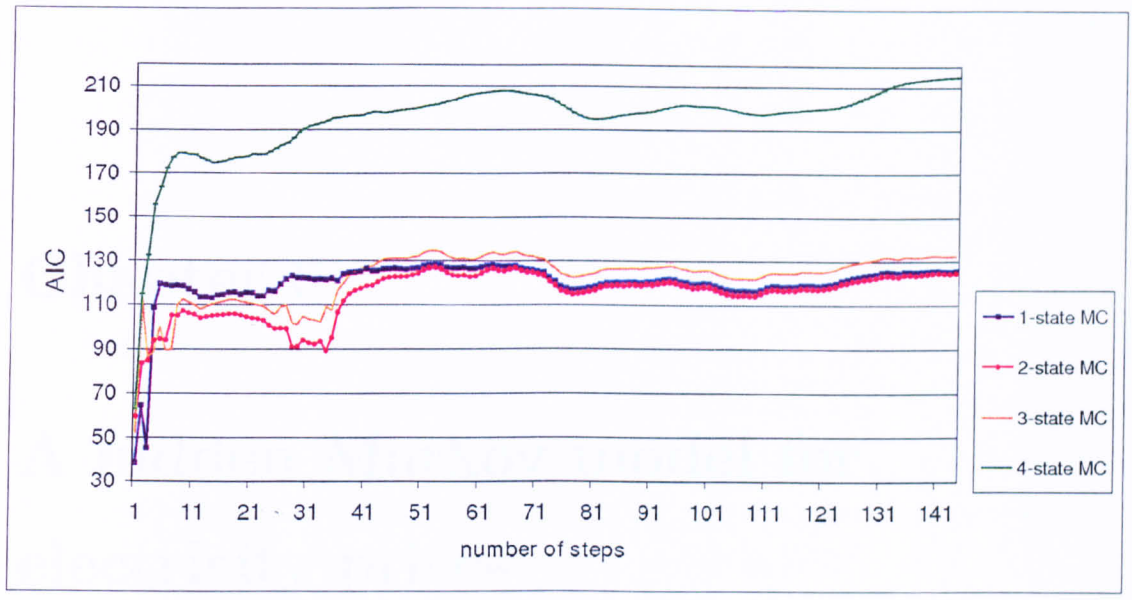


Figure 5.6: AIC for the 1-, 2-, 3- and 4-state HMM-based interest rate model

5.5 Some concluding remarks

In this chapter, a mean-reverting interest rate model with Markov-modulated parameters is developed and implemented. HMM filtering techniques detailed in the previous chapter are employed to obtain optimal estimates of the model parameters via recursive filters of auxiliary quantities of the observation process. Algorithms were designed and applied to a financial data set of 30-day Canadian T-bill yields. Within the data set and period studied, the conducted analysis shows that a model with two regimes is sufficient to describe the interest rate dynamics on the basis of very small prediction errors and the Akaike information criterion.

Chapter 6

A hidden Markov model for electricity prices

In this chapter a regime-switching model for electricity spot price dynamics is developed. The spot price is assumed to follow an exponential Ornstein-Uhlenbeck process with an added compound Poisson process. This way, the model allows for mean-reversion and possible jumps. All parameters are modulated by a hidden Markov chain in discrete time and are able to switch between different economic regimes representing the interaction of various factors. For this model the filtering method from chapter 4 is extended. With a different reference probability measure including the jump term optimal estimates of the model parameters can be derived in terms of the recursive filters. This self-calibrating model is implemented on a deseasonalised series of daily spot electricity prices from the Nordic exchange Nord Pool.

6.1 Modelling electricity prices

Electricity markets in many countries such as Norway, Spain, Germany, the UK and the US, amongst others, were deregulated over the last two decades. Price controls were removed and the competition between electricity providers encouraged. Electricity spot and future contracts are now open for trade and with this, the price uncertainty has increased. New models for pricing these electricity contracts were

introduced, which have to take into account distinctive stochastic properties of the electricity market.

Trading electricity products in spot and derivative markets largely resembles the trade with other commodities. However, electricity is an economical non-storable good, which leads to a strong dependency on supply and demand. High seasonal variations in prices arise from seasonal demand fluctuations, which occur weekly with different demands for weekdays and weekends as well as in an annual cycle due to changes in temperatures and daylight hours. The second characteristic of electricity prices is mean-reversion. Since electricity prices depend on supply and demand, the economic principle applies here; higher demand leads to higher prices, but in the long run the supply side will adapt to higher demand and prices will decrease to the long-run mean. The dependence of prices on supply and demand and the non-storability of electricity leads to the third characteristic of electricity prices, namely high price spikes. Supply shortages together with higher demand result in price jumps on the electricity spot market, which were frequently empirically observed over the last years. The challenge of a pricing model on the electricity market is therefore to be flexible enough to take all these characteristics into account.

The stochastic properties of electricity prices have led to different modelling approaches in the literature, which can be divided into two categories, namely forward based models and modelling spot prices. The first approach was taken by Clewlow and Strickland [34], Benth and Koekebakker [17], and Kiesel, Schindlmayer and Börger [93], who choose to model the entire forward curve directly to price forward and futures contracts. The second approach, which tends to be more tractable, was taken by Schwartz [134], where an OU-process modelled the mean-reversion. This approach was extended by Lucia and Schwartz [105], who proposed a two-factor mean-reverting model for spot prices with a deterministic component for the seasonal pattern. These models capture the mean-reversion but do not take into account the occurring price spikes. A mean-reverting jump-diffusion model for hourly spot prices was proposed by Culot et al. [36] and Cartea and Figueroa [31]. Another approach was taken by Benth et al. [16], where the characteristics

of spot prices are captured with mean-reversion dynamics driven by Lévy (jump) processes. Geman and Roncoroni [70] introduce a class of discontinuous processes to capture trajectorial and statistical properties of electricity prices. The model is fitted to the US market, where it captures specific jump characteristics.

In recent years, regime-switching models for electricity prices emerged. A study by de Jong [38] found that spikes in spot electricity prices can be captured better by regime-switching models than by a Poisson jump model. Deng [43] and de Jong and Huisman [39] developed different regime-switching models for electricity prices. Most regime-switching models distinguish between two regimes, one 'normal' and one 'jump' regime. Huisman and Mahieu [90] introduced a third regime for the change from 'jump' to 'normal' regime. However, the calibration and parameter estimation in all these models remains problematic due to limited historical data and a large number of parameters.

The model proposed in this chapter is an HMM, which is a more general version of regime-switching models. Elliott, Sick and Stein [57] introduced a Markov model for electricity spot prices including factors from the supply side in their analysis. The number of generators on-line are represented by a Markov process in discrete time and parameters are estimated using the EM algorithm. The concept of HMMs was applied to electricity markets by Yu and Sheblé [148] describing the structure of the electricity market with an HMM, and by González et al. [74], who use an Input/Output HMM for analysing electricity prices. Here, an HMM for forecasting electricity spot prices is developed, which incorporates seasonality, mean-reversion as well as the possibility of price spikes. The electricity spot price is the observation process of the HMM. Various information, which are difficult to quantify, such as behavioural aspects of buyers and sellers, unforeseen weather, and production issues, amongst others, are hidden in this process. This hidden information is modelled by a Markov chain in discrete time and governs the parameters of the proposed model. One main problem in forecasting prices in the electricity market is the estimation of parameters since daily prices can be very volatile and jumps can occur throughout the year. This HMM is a mean-reverting model with jumps, where the parameters evolve according to the underlying discrete-time

Markov chain.

In this chapter, recursive formulae for optimal model parameter estimation are derived through the reference measure technique by Elliott [48] and by applying the filters derived in chapter 4, namely equations (4.7), (4.9) and (4.11). In this new model, the OU-process from chapter 5 is extended by an addition of a jump-component. A new reference measure is constructed and the optimal parameter estimation includes the jump-component. The recursive parameter updates are implemented on a data set from Nordic exchange Nord Pool. The Nordic electricity market is known for its volatile nature. Since power production relies heavily on hydro power, water levels in the reservoirs are an important determinant in electricity generation and prices in the Nordic market have high seasonal variations and spikes. The forecasts in a 2- and 3-state HMM setting follow the actual data closely. Furthermore within the framework of this model, expected spots on delivery are calculated, which in turn can be incorporated in forward pricing for practitioners to adopt.

6.2 Model description

Consider the underlying probability space $(\Omega, \mathcal{F}, P)$. The electricity spot price model is composed of two main components: one deterministic function $D(k)$ to capture seasonal trends, and Z, an OU-process with added compound Poisson process, where the parameters are governed by a Markov chain. The OU-process part of the observation process models the mean-reversion of electricity prices observed in the market. The random price fluctuations are modelled by a Brownian motion W to include the 'normal' variations when the market is quiet. The compound poisson process Y models the price spikes.

As in chapters 3 and 5 the homogeneous Markov chain $\mathbf{x}_k$ governing the parameters of the model is defined in discrete time $(k = 0, 1, ...)$ with finite state space $\{\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_N}\}$. Since electricity spot prices are highly dependent on supply and demand, different states of the Markov chain represent different regimes of the market which are determined by the interaction amongst many variables such as the

current state of supply and demand and strategic behaviour of market agents. An exact definition of the elements of the hidden states is not necessary, all hidden information is included in the different states. The filtration generated by the Markov chain is denoted by $\mathcal{F}^{\mathbf{x}}$. Under the probability measure P the Markov chain $\mathbf{x}$ has dynamics $\mathbf{x}_{k+1} = \mathbf{\Pi}\mathbf{x}_k + \boldsymbol{v}_{k+1}$, where $\mathbf{\Pi}$ denotes the transition probability matrix of the Markov chain and $\boldsymbol{v}_{k+1}$ is a martingale increment with $\mathrm{E}[\boldsymbol{v}_{k+1} \mid \mathcal{F}_k^{\mathbf{x}}] = \mathbf{0}$. The spot price dynamics of daily electricity spot prices is given by

$$SP(k) = D(k)\exp(Z_k). \tag{6.1}$$

6.2.1 Seasonal decomposition

The seasonal component $D(k)$ is modelled with a sinusoidal function with positive trend. Sinusoidal functions are applied in various studies for capturing seasonal components. A sinusoidal seasonal function with weekly, semi-weekly and annual components was used by Culot et al. [36] and Kluge [95]. Furthermore, de Jong [38] introduced sinusoidal terms as well as dummy variables within his regime-switching approach, whilst Geman and Roncoroni [70] include a cosine function in the mean trend for the US market. Here, a sinusoidal function that has yearly and weekly components is chosen because the electricity demand on the Nordic market shows seasonal patterns for colder and warmer times of the year as well as weekly seasonal patterns. The seasonal component is given by

$$\begin{aligned} D(k) \;=\; & d1 * k + \sum_{h=1}^{3}\Bigg(d2_h \sin\Big(s_h\frac{2\pi}{365}k\Big) + d3_h \cos\Big(s_h\frac{2\pi}{365}k\Big) \\ & + d4_h \sin\Big(s_h\frac{2\pi}{7}k\Big) + d5_h \cos\Big(s_h\frac{2\pi}{7}k\Big)\Bigg) + d6 \end{aligned} \tag{6.2}$$

for $s_1 = 1, s_2 = 2$ and $s_3 = 4$ and the constants $d1$,$d2_h$ $d3_h$, $d4_h$, $d5_h$ and $d6$ are to be determined. This seasonal function is fitted to the data set in section 6.5.1.

6.2.2 Dynamics of the observation process

The deseasonalised log spot price $\ln \frac{SP(k)}{D(k)}$ is modelled through the stochastic process Z_k, a discretised version of Z_t, which is an OU-process with added jump component.

The process Z_t follows the dynamics

$$dZ_t = \kappa(\mathbf{x}_t)(\beta(\mathbf{x}_t) - Z_t)dt + \zeta(\mathbf{x}_t)\, dW_t + dY_t \tag{6.3}$$

where the mean-reversion level β, speed of mean-reversion κ and volatility ζ are all governed by the Markov chain $\mathbf{x}_t$. Since $\mathbf{x}_t$ is any one of the unit vectors $\mathbf{e}_i$, the parameters κ, β and ζ have the representations $\langle \cdot, \cdot \rangle$ with $\kappa(\mathbf{x}_t) = \langle \boldsymbol{\kappa}, \mathbf{x}_t \rangle$, $\beta(\mathbf{x}_t) = \langle \boldsymbol{\beta}, \mathbf{x}_t \rangle$ and $\zeta(\mathbf{x}_t) = \langle \boldsymbol{\zeta}, \mathbf{x}_t \rangle$. The jump process Y_t is given by

$$dY_t = B dG_t, \tag{6.4}$$

where G_t is a Poisson process with constant intensity λ^P. When a jump occurs, its magnitude B is a random variable with a probability density function that depends on the Markov chain $\mathbf{x}_t$ meaning that different regimes have different jump size distributions. The conditional distribution of the jump sizes is $B|\mathbf{x}_t \sim N(\mu_B(\mathbf{x}_t), \zeta_B^2(\mathbf{x}_t))$. The intensity λ^P does not change when a switching of regimes occurs. The seasonality of jump intensity is still taken into account, since the jump size is evolving according to the state of the Markov chain. The global filtration is defined by $\mathcal{F} = \mathcal{F}^W \vee \mathcal{F}^Y \vee \mathcal{F}^{\mathbf{x}}$, where the filtration generated by the Brownian motion is denoted by $\mathcal{F}^W$ and the filtration generated by the jump process component is $\mathcal{F}^Y$. The global filtration includes the filtration generated by the observation process $\mathcal{F}^Z = \sigma(Z_1, Z_2, ...)$.

6.3 Change of measure

For the derivation of optimal parameter estimates within this extended HMM set-up the reference probability measure technique as described in section 2.3.1 is applied. First the observation process is defined as the logarithm of the deseasonalised electricity spot prices. The continuous-time solution of equation (6.3) is

$$\begin{aligned} Z_t &= \ln \frac{SP(t)}{D(t)} \\ &= Z_s e^{-\kappa(\mathbf{x}_s)(t-s)} + \beta(\mathbf{x}_s)(1 - e^{-\kappa(\mathbf{x}_s)(t-s)}) \\ &\quad + \zeta(\mathbf{x}_s) e^{-\kappa(\mathbf{x}_s)t} \int_s^t e^{\kappa(\mathbf{x}_s)u}\, dW_u + \sum_{m=G_s+1}^{G_t} e^{-\kappa(\mathbf{x}_s)(t-\tau_m)} B_m(\mathbf{x}_t). \end{aligned} \tag{6.5}$$

The parameters $\boldsymbol{\kappa}, \boldsymbol{\beta}, \boldsymbol{\zeta}$ and the jump-size $\boldsymbol{B}_t$ of the compound Poisson process component are governed by a Markov chain $\mathbf{x}_t$ in discrete time. The Markov chain is constant over the interval $[s,t]$. The random time of occurrence of the mth jump is denoted by τ_m. For the derivation of the filters of related processes of the Markov chain $\mathbf{x}$ and in finding optimal parameter estimates we work under a reference probability measure $\bar{P}$. The discrete-time version of the observation process (6.5) is given by

$$\begin{aligned} Z_{k+1} &= Z_k e^{-\kappa(\mathbf{x}_k)\Delta k} + \beta(\mathbf{x}_k)(1 - e^{-\kappa(\mathbf{x}_k)\Delta k}) \\ &\quad + \zeta(\mathbf{x}_k)\sqrt{\frac{1 - e^{-2\kappa(\mathbf{x}_k)\Delta k}}{2\kappa(\mathbf{x}_k)}} w_{k+1} + \sum_{m=1}^{G_{\Delta k}} e^{-\kappa(\mathbf{x}_k)(\Delta k - \tau_m)} B_m(\mathbf{x}_k) \end{aligned} \tag{6.6}$$

where $\{w_{k+1}\}$ is a sequence of IID standard normal random variables. Note the following connection for the discretisation of the jump-term:

$$\begin{aligned} \int_l^k e^{-\kappa(\mathbf{x}_k)(k-u)} dY_u &= \sum_{m=G_k+1}^{G_{k+1}} e^{-\kappa(\mathbf{x}_k)(k-\tau_m)} B_m(\mathbf{x}_k) \\ &\overset{\text{in distr.}}{=} e^{-\kappa(\mathbf{x}_k)(k-l)} \sum_{m=1}^{G_{k-l}} e^{\kappa(\mathbf{x}_k)\tau_m} B_m(\mathbf{x}_k), \end{aligned}$$

where τ_m are the jumping times in the interval $(0, k-l]$.

The filters for the related processes of the Markov chain (4.7), (4.9) and (4.11) are calculated under a reference probability measure $\bar{P}$. Under this measure $\mathbf{x}$ is still a Markov chain with dynamics $\mathbf{x}_{k+1} = \boldsymbol{\Pi}\mathbf{x}_k + \boldsymbol{v}_{k+1}$ and Z_k are independent observations. To perform a change of measure we examine the discretised observation process. We assume that the change of measure does not affect the compound Poisson process component of the observation process as in Merton [116], where the jump size and intensity have the same dynamics under the new measure. We construct a reference probability measure $\bar{P}$ through the Radon-Nikodým derivative $\frac{d\bar{P}}{dP} \mid_{\mathcal{F}_k} = \Lambda_k = \prod_{l=1}^{k} \lambda_l$. Following the technique developed in Elliott et al. [49] we back out the real world measure from the reference probability measure by applying a discrete time version of Girsanov's theorem and defining

$\frac{dP}{d\bar{P}}\big|_{\mathcal{F}_k} = \bar{\Lambda}_k = \prod_{l=1}^{k} \bar{\lambda}_l$ with

$$\begin{aligned}\bar{\lambda}_l \;:=\; \exp\Big(&-\frac{1}{b}\big[Z_l e^{-\kappa(\mathbf{x}_l)\Delta l} + \beta(\mathbf{x}_l)(1-e^{-\kappa(\mathbf{x}_l)\Delta l}) \\ &+ \sum_{m=1}^{G_{\Delta l}} e^{-\kappa(\mathbf{x}_l)(\Delta l-\tau_m)} B_m(\mathbf{x}_l)\big]\frac{Z_{l+1}}{b} \\ &- \frac{1}{2b^2}\big[Z_l e^{-\kappa(\mathbf{x}_l)\Delta l} + \beta(\mathbf{x}_l)(1-e^{-\kappa(\mathbf{x}_l)\Delta l}) \\ &+ \sum_{m=1}^{G_{\Delta l}} e^{-\kappa(\mathbf{x}_l)(\Delta l-\tau_m)} B_m(\mathbf{x}_l)\big]^2\Big) \qquad (6.7)\end{aligned}$$

with $b = \zeta(\mathbf{x}_l)\sqrt{\frac{1-e^{-2\kappa(\mathbf{x}_l)\Delta l}}{2\kappa(\mathbf{x}_l)}}$ and $\bar{\Lambda}_0 = 1, \{\bar{\lambda}_l : l \in \mathbb{N}\}$ and $\{\bar{\Lambda}_l : l \in \mathbb{N}\}$. The process $\{\bar{\Lambda}_l\}$ is a $\mathcal{F}$-martingale under $\bar{P}$ and $\bar{\Lambda}\Lambda = 1$ with

$$\begin{aligned}\Lambda_k \;=\; \prod_{l=1}^{k}\Bigg(&\exp\Big[\frac{1}{b}\big[Z_l e^{-\kappa(\mathbf{x}_l)\Delta l} + \beta(\mathbf{x}_l)(1-e^{-\kappa(\mathbf{x}_l)\Delta l}) + \sum_{m=1}^{G_{\Delta l}} e^{-\kappa(\mathbf{x}_l)(\Delta l-\tau_m)} B_m(\mathbf{x}_l)\big] w_{l+1} \\ &+\frac{3}{2b^2}\big[Z_l e^{-\kappa(\mathbf{x}_l)\Delta l} + \beta(\mathbf{x}_l)(1-e^{-\kappa(\mathbf{x}_l)\Delta l}) + \sum_{m=1}^{G_{\Delta l}} e^{-\kappa(\mathbf{x}_l)(\Delta l-\tau_m)} B_m(\mathbf{x}_l)\big]^2\Big]\Bigg).\end{aligned}$$

Note here, that the reference probability measure is not unique, other choices for $\bar{\Lambda}_k$ are possible. Recall from chapter 4 that with Bayes' theorem, a filter for an adapted process H_k is given by

$$\mathrm{E}\left[H_k \mid \mathcal{F}_k^{\mathbf{x}}\right] = \frac{\bar{\mathrm{E}}\left[H_k\bar{\Lambda}_k \mid \mathcal{F}_k^{\mathbf{x}}\right]}{\bar{\mathrm{E}}\left[\bar{\Lambda}_k \mid \mathcal{F}_k^{\mathbf{x}}\right]}$$

and has the representation $\mathrm{E}\left[H_k \mid \mathcal{F}_k^{\mathbf{x}}\right] = \dfrac{\langle \mathbf{1}, \eta_k(H_k\mathbf{x}_k)\rangle}{\langle \mathbf{1}, \eta_k(\mathbf{x}_k)\rangle}$.

The filters for the state of the Markov chain, the number of jumps J, the occupation time O and auxiliary processes T are applied, which were derived in chapter 4. From the change of measure described above, Γ^i in the filter equations (4.7), (4.9) and (4.11) is defined by

$$\begin{aligned}\Gamma^i \;=\; \exp\Big(&-\frac{1}{b_i^2}\big[Z_l e^{-\kappa_i\Delta l} + \beta_i(1-e^{-\kappa_i\Delta l}) + \sum_{m=1}^{G_{\Delta l}} e^{-\kappa_i(\Delta l-\tau_m)} B_m^i\big] Z_{l+1} \\ &- \frac{1}{2b_i^2}\big[Z_l e^{-\kappa_i\Delta l} + \beta_i(1-e^{-\kappa_i\Delta l}) + \sum_{m=1}^{G_{\Delta l}} e^{-\kappa_i(\Delta l-\tau_m)} B_m^i\big]^2\Big). \qquad (6.8)\end{aligned}$$

6.4 Optimal parameter estimates

In this section, maximum likelihood estimates for the parameters of the observation process Z_t in equation (6.5) are derived, where the parameters are governed by a Markov chain $\mathbf{x}_t$.

First we derive the probability density function of the process Z_t.

$$\begin{aligned} Z_t &= Z_s e^{-\kappa(\mathbf{x}_s)(t-s)} + \beta(\mathbf{x}_s)(1 - e^{-\kappa(\mathbf{x}_s)(t-s)}) \\ &\quad + \zeta(\mathbf{x}_s)e^{-\kappa(\mathbf{x}_s)t} \int_s^t e^{\kappa(\mathbf{x}_s)u}\, dW_u + \sum_{m=G_s+1}^{G_t} e^{-\kappa(\mathbf{x}_s)(t-\tau_m)} B_m(\mathbf{x}_t). \end{aligned} \tag{6.9}$$

The parameters are said to be constant over the interval $[s,t], 0 \leq s \leq t$. The observation process without jumps is normally distributed with mean $\mu_Z = \beta + (Z_s - \beta)e^{-\kappa(t-s)}$ and variance $\zeta_Z^2 = \frac{\zeta^2}{2\kappa}(1 - e^{-2\kappa(t-s)})$. We examine the distribution of the part given by the compound Poisson process Y_t. As described in the previous section $B_1, B_2, \ldots$ are independent, identically distributed normal random variables and $(G_t)_{t\geq 0}$ is a standard Poisson process with jump intensity $\lambda^P > 0$. Let G and B be jointly independent. Denote the mean and the variance of the process B by μ_B and ζ_B^2, respectively. The probability distribution of the Poisson process G is given by the usual Poisson distribution. To derive the density of the jump component, the approximation of the jump integral $\int_s^t e^{-\kappa(t-u)} dY_u \approx e^{-\kappa(t-s)}(Y_t - Y_s)$ is utilised. By the stationarity of the compound Poisson process, we find that the increment $Y_t - Y_s$ has the same distribution as Y_{t-s}. Thus, the density of the the jump term is

$$\Phi_{Y_{t-s}}(z) = \sum_{h=0}^{\infty} \frac{(\lambda^P(t-s))^h}{h!} e^{-\lambda^P(t-s)}\, \phi(z; \mu_B e^{-\kappa(t-s)}h, \zeta_B^2 e^{-2\kappa(t-s)}h)\,, \tag{6.10}$$

where ϕ denotes the pdf of the normal distribution. Following the arguments by Hanson and Westman [84] the pdf of the observation process can be calculated as the convolution of densities of the OU-process without jumps and the jump part

distribution. Therefore the density of Z_t conditioned on Z_s is

$$\Phi_Z(x) = \sum_{h=0}^{\infty} \frac{(\lambda^P(t-s))^h}{h!} e^{-\lambda^P(t-s)} \times \phi\Big(z; \beta + (Z_s - \beta)e^{-\kappa(t-s)} + \mu_B e^{-\kappa(t-s)} h, \frac{\zeta^2}{2\kappa}(1 - e^{-2\kappa(t-s)}) + \zeta_B^2 e^{-2\kappa(t-s)} h\Big). \tag{6.11}$$

The density in (6.11) can be further expressed as an expectation of the normal density under the Poisson counter $G_{\Delta t}$. Hence, equation (6.11) has a more compact form

$$\begin{aligned} \Phi_X(x) &= \mathrm{E}_{G_{\Delta k}}\Big[\phi(x; \beta + (Z_s - \beta)e^{-\kappa(t-s)} + \mu_B e^{-\kappa(t-s)} G_{\Delta t}, \\ &\qquad \frac{\zeta^2}{2\kappa}(1 - e^{-2\kappa(t-s)}) + \zeta_B^2 e^{-2\kappa(t-s)} G_{\Delta t})\Big]. \end{aligned} \tag{6.12}$$

To find the optimal parameters of the observation process Z_t specified in (6.5) the EM algorithm is utilised. For this purpose the simplifying assumption is made, that the intensity of the Poisson process λ^P is independent from the other parameters. To find optimal estimates, the parameters of the normally distributed part of the observation process are evaluated independently of the G_t-process. First, the maximum likelihood estimates for the set of parameters $\rho_{\text{elec}} = \{\kappa_i, \beta_i, \zeta_i^2, \mu_{B_i}, \zeta_{B_i}^2, \pi_{ji}\}$ is derived. The aim is to find a new set of parameters $\hat{\rho}_{\text{elec}}$, which maximises the conditional expectation of the log-likelihoods. In the following we denote the jump counter $G_{\Delta k}$ with p and the mean and variance of the OU-process with μ_Z and ζ_Z, respectively. The MLE's for the normal distribution $\phi(x; \mu_Z + \mu_B p e^{-\kappa(t-s)}, \zeta_Z^2 + \zeta_B^2 p e^{-\kappa(t-s)})$ are derived. Note that both mean and variance are dependent on the Markov chain $\mathbf{x}$, and so they are regime-switching. The discretised version of the observation process in equation (6.6) is used in obtaining the recursive parameter updates.

An explicit recursive formula for the parameter β with the processes of the Markov chain $\mathbf{x}$ is derived. However, since the mean-reversion level κ is included in the mean and variance part, the calculation of the MLE for κ is less straightforward and a recursive formula cannot be found. Therefore explicit recursive formulae for the mean

μ_Z and the mean of the jump process μ_B are derived and a value for κ is obtained based on the optimal value of β by solving the equation $\mu_{Z_i} = \beta_i + (Z_l - \beta_i)e^{-\kappa_i \Delta}$. Consequently $\kappa_i = -\ln\left(\frac{\mu_{Z_i} - \beta_i}{X_l - \beta_i}\right)\frac{1}{\Delta}$. With this value of κ together with the MLE estimate of ζ_Z the estimated value of ζ_i is given by $\zeta_i^2 = \frac{2\kappa_i \zeta_{Z_i}^2}{1 - e^{-2\kappa_i \Delta}}$. Therefore, calculating MLE's for μ_Z and ζ_Z gives the desired parameter estimates for κ and ζ^2. Applying the EM algorithm the following optimal recursive parameter estimates are obtained:

$$\hat{\mu}_{Z_i} = \frac{\hat{T}_k^i(Z_{k+1}) - \hat{O}_k^i \mu_{B_i} p e^{-\kappa_i \Delta}}{\hat{O}_k^i} \tag{6.13}$$

$$\hat{\mu}_{B_i} = \frac{\hat{T}_k^i(Z_{k+1}) - \hat{O}_k^i \mu_{Z_i}}{\hat{O}_k^i p e^{-\kappa_i \Delta}} \tag{6.14}$$

$$\begin{aligned} \hat{\beta}_i = & \frac{\hat{T}_k^i(Z_l)(e^{-2\kappa_i \Delta} + e^{-\kappa_i \Delta}) - \hat{O}_k^i(-e^{-2\kappa_i \Delta}\mu_{B_i} p + \mu_{B_i} p e^{-\kappa_i \Delta})}{\hat{O}_k^i(1 + 2e^{-\kappa_i \Delta} + e^{-2\kappa_i \Delta})} \\ & + \frac{\hat{T}_k^i(Z_{l+1})(1 - e^{-\kappa_i \Delta})}{\hat{O}_k^i(1 + 2e^{-\kappa_i \Delta} + e^{-2\kappa_i \Delta})} \end{aligned} \tag{6.15}$$

$$\begin{aligned} \hat{\zeta}_{Z_i}^2 = & \frac{\hat{T}_k^i(Z_{k+1}^2) + \hat{O}_k^i(\mu_{Z_i}^2 + \mu_{B_i}^2 p^2 e^{-2\kappa_i \Delta}) + \hat{O}_k^i(2\mu_{Z_i}\mu_{B_i} p e^{-\kappa_i \Delta} - \zeta_{B_i}^2 e^{-2\kappa_i \Delta} p)}{\hat{O}_k^i} \\ & - \frac{2\hat{T}_k^i(Z_{k+1})(\mu_{Z_i} + \mu_{B_i} p e^{-\kappa_i \Delta})}{\hat{O}_k^i} \end{aligned} \tag{6.16}$$

$$\begin{aligned} \hat{\zeta}_{B_i}^2 = & \frac{\hat{T}_k^i(Z_{k+1}^2) + \hat{O}_k^i(\mu_{Z_i}^2 + \mu_{B_i}^2 p^2 e^{-2\kappa_i \Delta}) + \hat{O}_k^i(2\mu_{Z_i}\mu_{B_i} p e^{-\kappa_i \Delta} - \zeta_{Z_i}^2)}{\hat{O}_k^i p e^{-2\kappa_i \Delta}} \\ & - \frac{2\hat{T}_k^i(Z_{k+1})(\mu_{Z_i} + \mu_{B_i} p e^{-\kappa_i \Delta})}{\hat{O}_k^i p e^{-2\kappa_i \Delta}} \end{aligned} \tag{6.17}$$

$$\text{and} \qquad \hat{\pi}_{ji} = \frac{\hat{J}_k^{ji}}{\hat{O}_k^i}. \tag{6.18}$$

Note that for any process H we write $\hat{H}_l = E[H_l \mid \mathcal{F}_k^X]$. The proofs of formulae (6.13)-(6.17) can be found in Appendices D.1 - D.5. The proof of formula (6.18) is similar to proof in Appendix B.2.

6.5 Implementation

The model is implemented on daily spot prices compiled by Nord Pool. Nord Pool is the Nordic Power Exchange for trading electric power and was established in 1993. It is the only multinational exchange for trading power, integrating the Nordic countries Norway, Sweden, Finland and Denmark. The data set in this implementation consists of 1360 data points for the time period between December 1998 and August 2002.

6.5.1 Fitting the deterministic function

Due to the nature of power production in Norway, where most electricity is produced from hydro power, electricity prices are highly dependent on water levels in reservoirs and snow melting conditions. Compared to more stable markets like that of Singapore or that of the US, the Nordic market shows more frequently occurring jumps and daily spot prices compiled by Nord Pool exhibit high seasonality. To remove seasonal components, the deterministic function is fitted to the actual data. The parameters for the deterministic function are calibrated with a least-square algorithm in Matlab. In particular, $\frac{1}{2}\sum_k (D(x,k) - SP(k))^2$ is minimised with respect to x, where x denotes a set of parameters. In this case, $x = \{d1, d2_h, d3_h, d4_h, d5_h, d6\}$, is a parameter set described in equation (6.2). The resulting best fitting deterministic function for the seasonality component has the estimated parameter values of $d1 = 0.0566$, $d2_1 = 14.2141$, $d2_2 = -0.6806$, $d2_3 = 4.6331$, $d3_1 = 20.4066$, $d3_2 = -1.6644$, $d3_3 = 2.6234$, $d4_1 = 8.5408$, $d4_2 = 3.1962$, $d4_3 = 0.7352$, $d5_1 = -0.5275$, $d5_2 = 4.3471$, $d5_3 = 2.8170$ and $d6 = 97.3887$.

In Figure 6.1 daily spot prices in NOK/MWh are depicted together with the seasonal function. Frequent jumps in the electricity prices are visible and the descriptive statistics show a high variance of the price data. The remaining stochastic part is the log of the deseasonalised spot price SP. This is considered as the observation process for the empirical work presented in the next subsection.

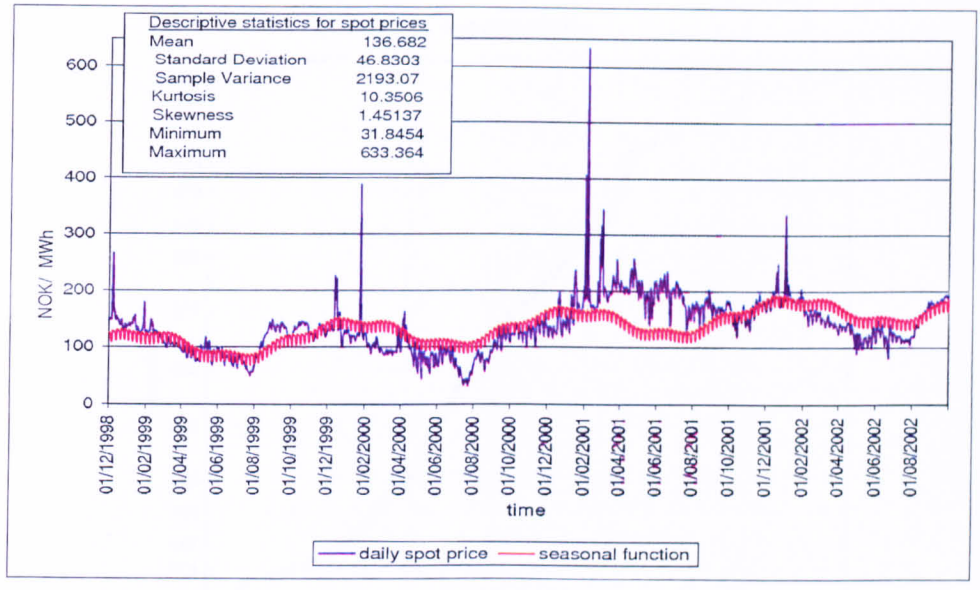


Figure 6.1: Actual data and seasonal function

6.5.2 Filtering and parameter estimation

First, a data analysis is conducted to verify that the dynamics for the stochastic process Z_k are suitable for this data set. A normality test for the deseasonalised log spot prices shows, that the spot prices do not follow a normal distribution. To test a possible normal distribution we perform the Jarque-Bera test at a 5%−level. With a p-value of 0, the hypothesis that the distribution is normal is rejected. Figure 6.2 shows a normality plot for the data set. The data deviates significantly from the straight line. This supports the model choice, the probability of rare events in the data sample is much higher than that predicted by a Gaussian distribution.

The filters for updating the model parameters are applied to the data set. Note that in the implementation hybrid states of the Markov chain are allowed. For instance, in a two-state setting $\mathbf{x}_k$ could take the value $(0.3, 0.7)^\top$, which indicates that $\mathbf{x}_k$ has a 0.3 probability of being in state 1 and a 0.7 probability of being in state 2. A series of one-step ahead forecasts for the deseasonalised daily log spot prices is calculated. The expected value of the observation process at time $k+1$ is

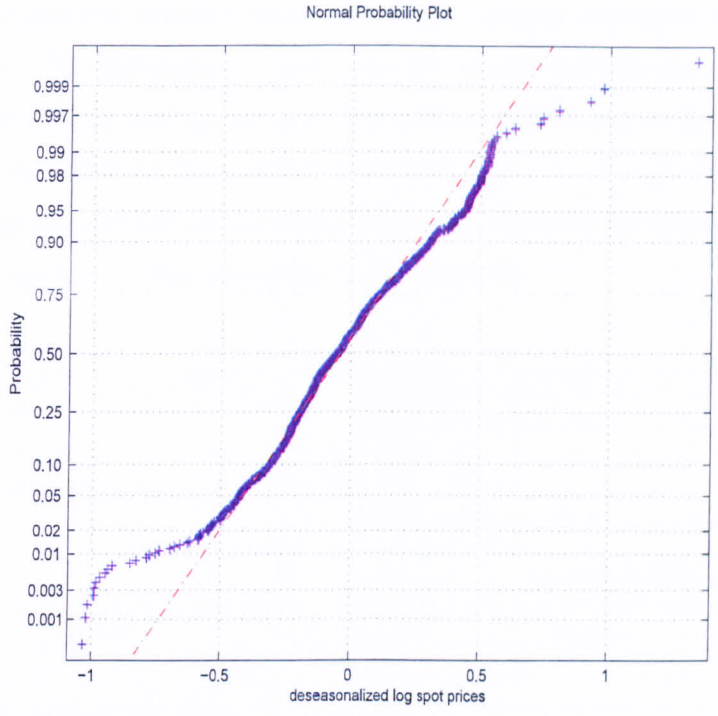


Figure 6.2: Normal probability plot for the deseasonalised log spot prices

obtained using the equation

$$\begin{aligned} \mathrm{E}[Z_{k+1} \mid \mathcal{F}_k] &= Z_k e^{-\langle \boldsymbol{\kappa}, \boldsymbol{\Pi}\mathbf{x}_k \rangle \Delta k} + \langle \boldsymbol{\beta}, \boldsymbol{\Pi}\mathbf{x}_k \rangle (1 - e^{-\langle \boldsymbol{\kappa}, \boldsymbol{\Pi}\mathbf{x}_k \rangle \Delta k}) \\ &\quad + \lambda^P k e^{-\langle \boldsymbol{\kappa}, \boldsymbol{\Pi}\mathbf{x}_k \rangle \Delta k} \langle \boldsymbol{\mu}_B, \boldsymbol{\Pi}\mathbf{x}_k \rangle \,. \end{aligned} \tag{6.19}$$

The mean value and the variance from the data set are used as guides to select initial values. Practitioners implementing the algorithm on different data sets may need to run the algorithm with different choices for the initial values as the EM algorithm only yields a local maximum for the log-likelihood. The data set is processed in batches of 40 data points. This means that roughly every six weeks (sufficient to warrant the presence of new information), the parameter estimates are updated based on updated filtering equations. The algorithm is run 34 times within this data set. The algorithm is self-tuning since new information leads to updated optimal parameter estimates. The implementation is performed under the set-up of a 2-state and 3-state Markov chain. To show the added benefit of introducing the switching of regimes, the algorithm is also run on a 1-state model setting, and the results are compared.

Figure 6.3 depicts the dynamic movement of the optimal parameter estimates of a 2-state HMM. The parameters κ and ζ are calculated through the updated optimal parameters β, μ_Z and ζ_Z. In updating κ the condition $\mu_{Z_i} > Z_l$ for $Z_l > \beta_i$ or $\mu_{Z_i} < Z_l$ for $Z_l < \beta_i$ must be satisfied, otherwise $\kappa_i(p+1) = \kappa_i(p)$ for $p = 1, ..., np$ with np being the number of parameter updates. All other parameters are calculated via the recursive filter equations (6.13) to (6.18) .

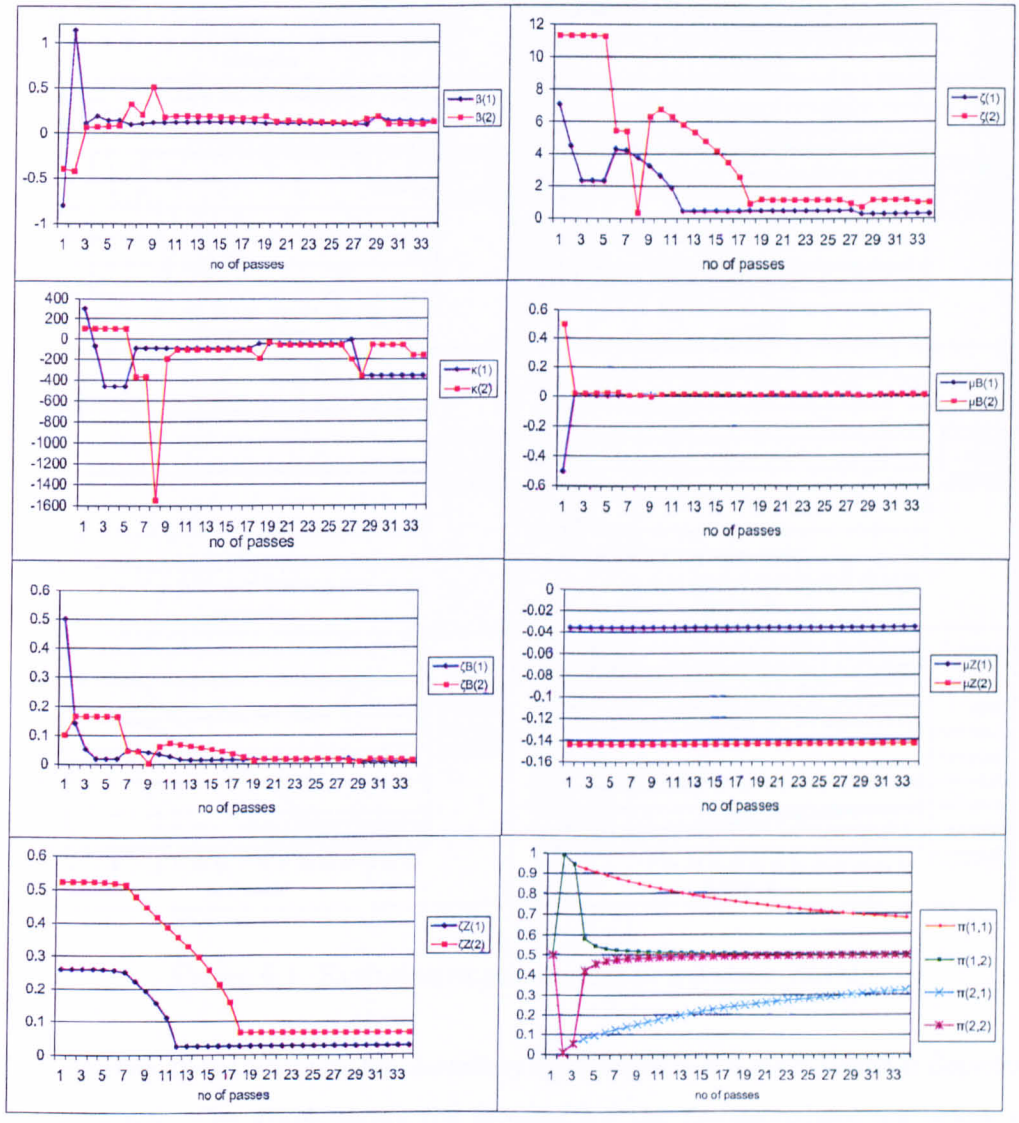


Figure 6.3: Evolution of parameters in a 2-state HMM

In Figure 6.4 the evolution of parameters in a 3-state Markov chain setting is displayed. Here, the evolution of parameters exhibits similar pattern to that of the

2-state Markov chain setting. The rate of convergence of parameter estimates is slightly higher than that in the 2-state set-up.

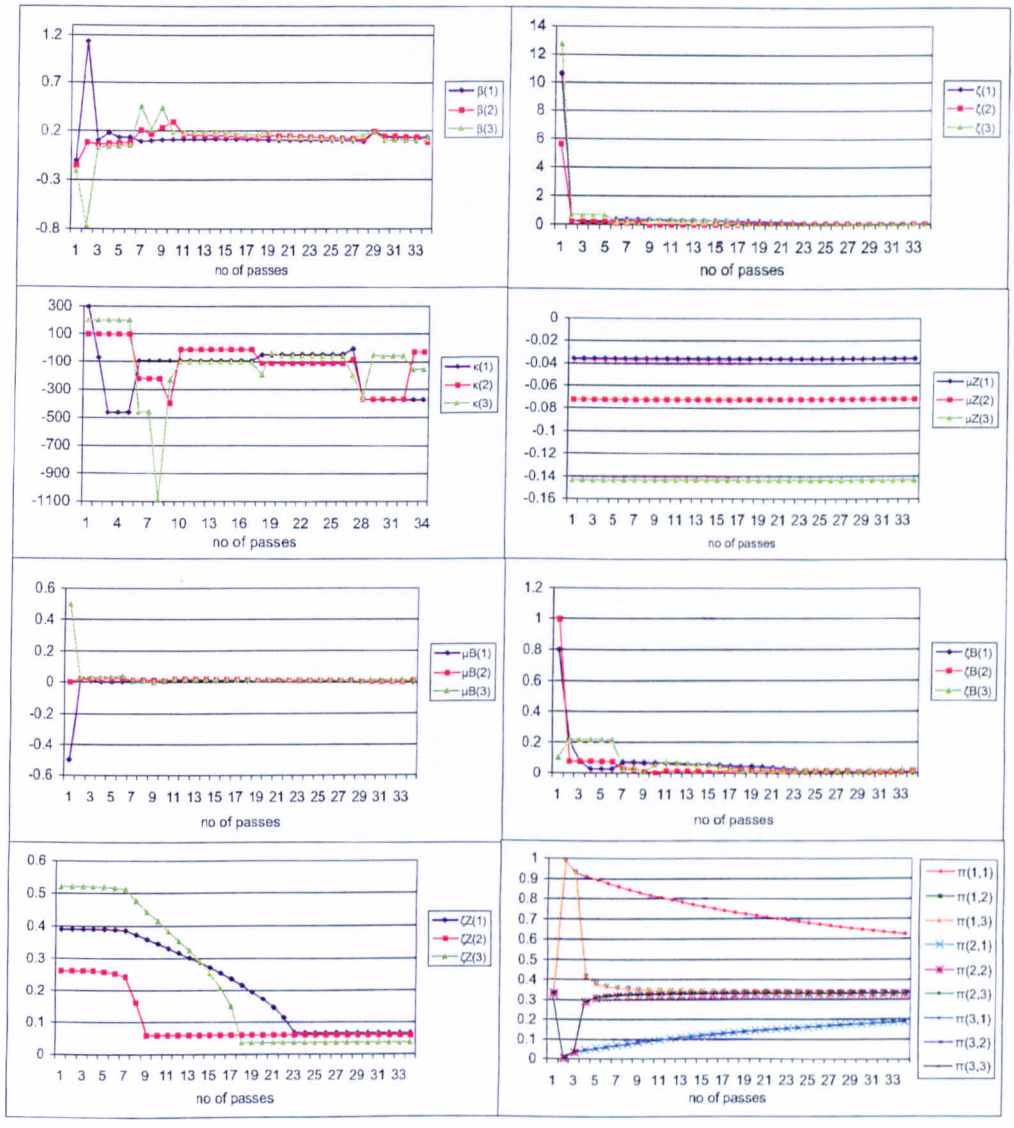


Figure 6.4: Evolution of parameters in a 3-state HMM

The one-step ahead forecasts for electricity spot prices in a 3-state HMM is depicted in Figure 6.5. The graph shows the forecast for the deseasonalised log prices as well as for the daily spot prices. Here we can see that the one-step ahead forecasts follow the actual values very closely. The self-tuning algorithm is able to capture the dynamics of the electricity spot prices and the occurrence of jumps is picked up by the filter.

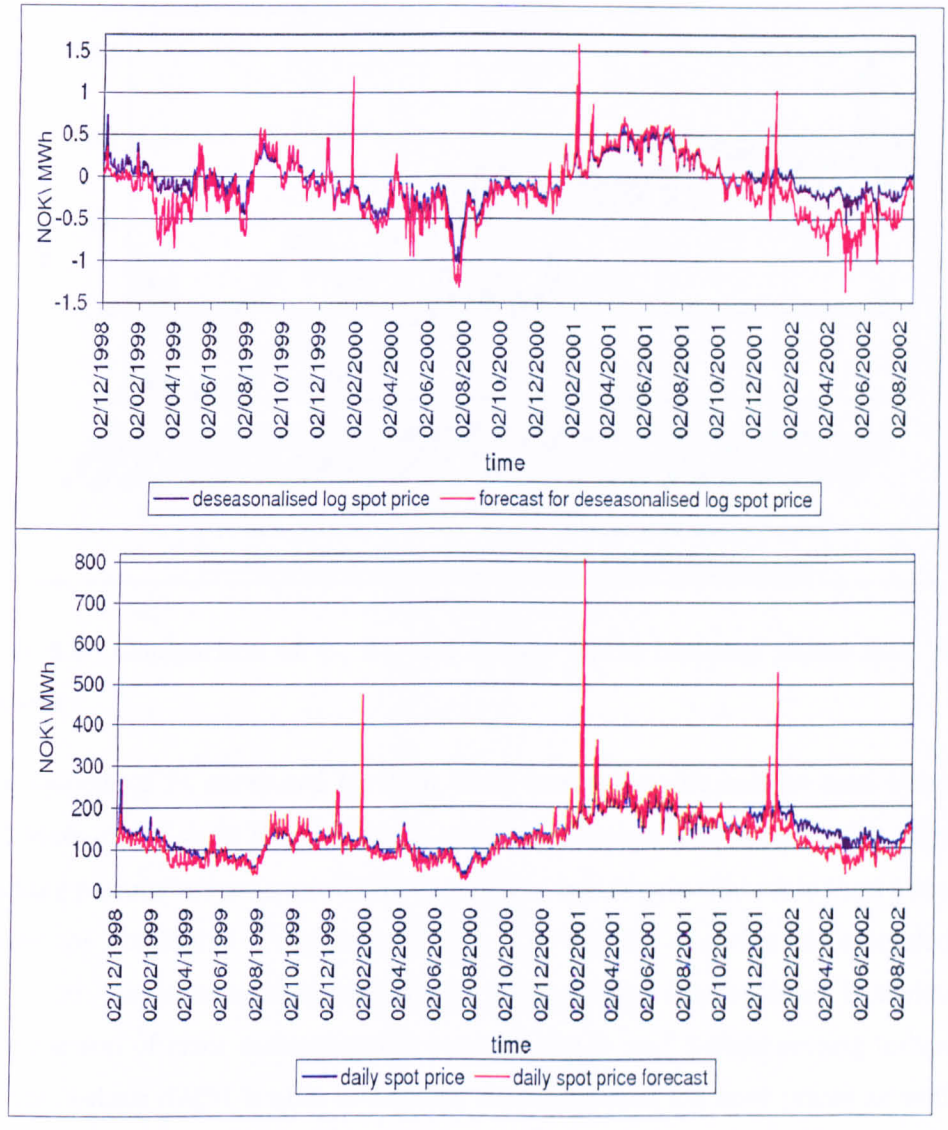


Figure 6.5: One-step ahead electricity price forecasts in a 3-state HMM

A comparison of the 1-, 2- and 3-state HMM forecasts shows that regime-switching significantly improves the one-step ahead forecasts. Figure 6.6 depicts a zoom-in view of forecasts from the three models during a 6-month period. The 2- and 3-state settings give closer forecasts than the 1-state model.

An error analysis on the deviations between forecasts and deseasonalised log spot prices as well as forecasts and daily spot prices is given in Table 6.1 utilising the error measures defined in subsection 3.6.2. The most substantial improvement in

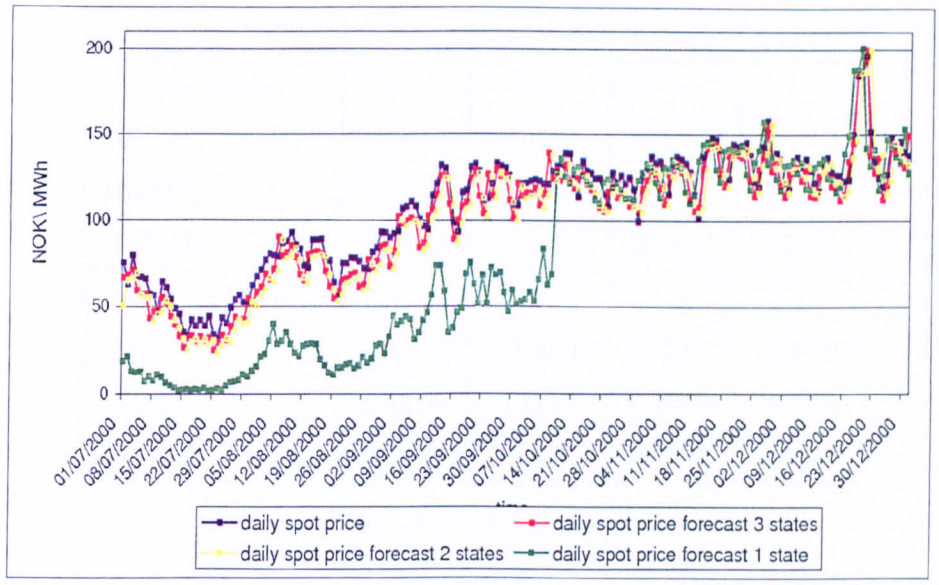


Figure 6.6: Comparison of 1-, 2-, and 3-state HMM one-step ahead spot price forecasts

the goodness of fit measured through these forecast errors can be seen between the 1-state and 2-state HMM. This is a clear indication that models with regime-switching parameters are more likely to generate better price forecasts than a model without the possibility of regime-switching. The number of states was extended to $N = 4$, but the error analysis shows that no significant improvement is achieved. A comparison of error measurements between the 2- and 3-state setting indicates that the 3-state HMM is able to forecast deseasonalised log spot prices as well as daily spot prices better than the 2-state model. Since an increase in the number of states adds to the complexity of parameter estimation, the slight improvement in terms of MdAPE using a 4-state setting does not really merit choosing a model with $N > 3$.

6.5.3 Pricing of electricity contracts

The optimal parameter estimates generated by the proposed filtering procedures are used in pricing the expected spot on delivery (ESD). The price which is considered here is given by the expected values of the spot price at a future time T given the current price at time t under the real-world measure P. The forward price is calculated as the ESD under a risk-neutral measure. The risk-premium can then

Error measure	1-state	2-state	3-state	4-state
log spot prices:				
MSE	0.4615	0.0577	0.0455	0.0464
MdAPE	2.066	0.7681	0.6771	0.6115
MdRAE	6.2049	3.3750	2.9837	2.7961
actual spot prices:				
MSE	4967.5	1160.8	1142.8	1026.9
MdAPE	0.2246	0.1283	0.1080	0.1046
MdRAE	0.9999	0.9677	0.9632	0.9627

Table 6.1: Error analysis for HMM forecasts

be calculated as the difference between these two prices.

The ESD in this model is calculated using

$$\begin{aligned}
ESD(SP_T, t, T) &= \mathrm{E}^P\big[SP_T \mid \mathcal{F}_t\big] \\
&= D(T)\exp(Z_t)^{e^{-\kappa(\mathbf{x}_t)(T-t)}} \\
&\quad \times \exp\Big[\beta(\mathbf{x}_t)\big(1-e^{-\kappa(\mathbf{x}_t)(T-t)}\big) + \frac{\zeta(\mathbf{x}_t)^2}{2}\big(1-e^{-2\kappa(\mathbf{x}_t)(T-t)}\big)\Big] \\
&\quad \times \mathrm{E}^P\big[\ \exp\big(\int_t^T e^{-\kappa(\mathbf{x}_t)(T-u)} B dG_u\big) \mid \mathcal{F}_t\big] \\
&= D(T)\Big[\frac{SP(t)}{D(t)}\Big]^{e^{-\kappa(\mathbf{x}_t)(T-t)}} \\
&\quad \times \exp\Big[\beta(\mathbf{x}_t)\big(1-e^{-\kappa(\mathbf{x}_t)(T-t)}\big) + \frac{\zeta(\mathbf{x}_t)^2}{2}\big(1-e^{-2\kappa(\mathbf{x}_t)(T-t)}\big)\Big] \\
&\quad \times \exp\Big[\int_t^T \exp\big(\mu_B(\mathbf{x}_t)e^{-\kappa(\mathbf{x}_t)(T-u)} \\
&\qquad + \frac{\zeta_B^2(\mathbf{x}_t)}{2}e^{-2\kappa(\mathbf{x}_t)(T-u)}\big)\lambda^P du - \lambda^P(T-t)\Big]. \qquad (6.20)
\end{aligned}$$

A proof for the expected value of an integral similar to the last term $\mathrm{E}^P[\exp(\int_t^T e^{-\kappa(\mathbf{x}_t)(T-u)} dY_u) \mid \mathcal{F}_t]$ is stated in Benth et al. [16].

The expectations above have to be calculated numerically if we do not assume that the Markov chain is constant. This approach is necessary for pricing ESD's with longer delivery periods which is an important issue for future research.

The ESD is computed for maturities of $T = 1, ..., 30$ days with starting dates from the 1st of July 2002 to the 15th of July 2002. Figures 6.7 and 6.8 show the ESD prices in a 2-state and 3-state setting, respectively. In this example for short delivery periods it is reasonable to assume a constant Markov chain. Since the starting dates lie in the last batch of data in our filtering estimation, we use the corresponding latest parameter estimates in a 2-state setting, which are $\kappa(\mathbf{x}_t) = -0.7731$, $\beta(\mathbf{x}_t) = 0.1327$, $\zeta(\mathbf{x}_t) = 0.55$, $\mu_B(\mathbf{x}_t) = 0.0075$ and $\zeta_B(\mathbf{x}_t) = 0.0079$ and the parameters in a 3-state setting, which are $\kappa(\mathbf{x}_t) = -0.5853$, $\beta(\mathbf{x}_t) = 0.1211$, $\zeta(\mathbf{x}_t) = 0.04$, $\mu_B(\mathbf{x}_t) = 0.0091$ and $\zeta_B(\mathbf{x}_t) = 0.0089$. The prices in a 2-state setting are slightly lower than the prices derived in a 3-state HMM; the difference increases with increasing maturity. This pricing example shows a practical application of the model for spot prices and the filtered parameters. For any deterministic market price of risk, the conclusion of this application may be carried over to forward prices.

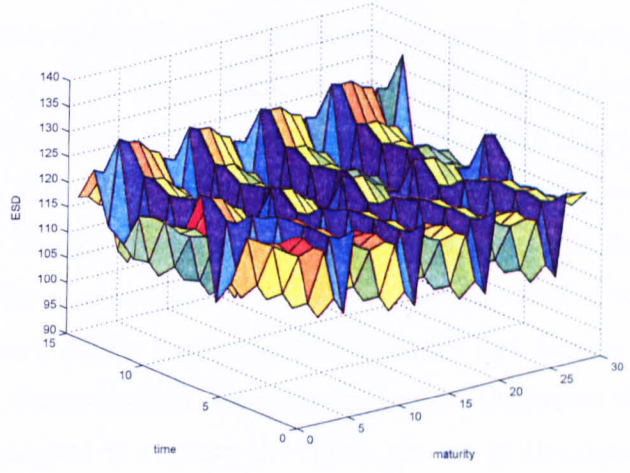


Figure 6.7: Expected spot on delivery in a 2-state HMM

6.6 Some concluding remarks

An HMM-driven model was developed to forecast electricity spot prices. The spot price is assumed to evolve in accordance with the exponential of an OU process plus a jump term and this exponential is scaled by a deterministic sinusoidal function to take into account the seasonal component of electricity prices. The added

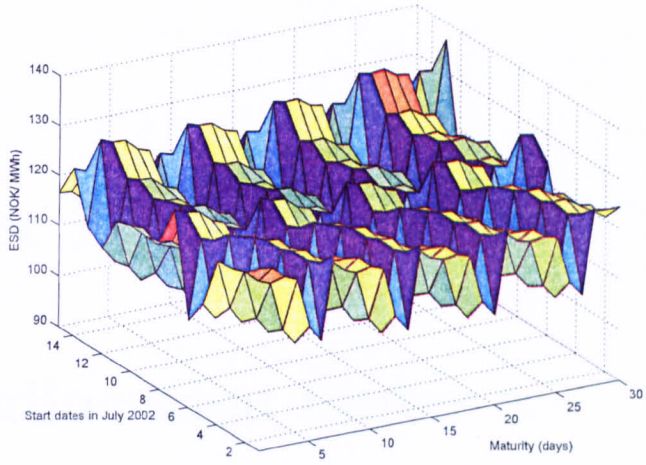


Figure 6.8: Expected spot on delivery in a 3-state HMM

compound Poisson process has normally distributed jumps, where the mean and variance are governed by a discrete-time HMM. This offers greater flexibility to switch between economic regimes reflected by the dynamic changes of various factors such as electricity supply and demand or behavioural aspects, which are easily seen in the sudden jumps of spot prices. Employing the EM algorithm, the optimal estimates for the model parameters are derived in terms of the recursive filters for the state of the Markov chain, the number of jumps between two states, occupation time of the Markov chain and an auxiliary process. Since the parameters are updated whenever a new data set arrives, we have created a self-tuning model. The empirical work on the implementation of filters and parameter estimation of the model using deseasonalised electricity spot prices illustrates that the proposed model is well-equipped to capture the spikes present in the data for both the 2-state and 3-state setting. A comparison with the 1-state model indicates clearly the improvements gained by allowing parameters to switch between regimes. The important empirically observed characteristics of the electricity markets are captured by the model as evidenced by low forecast errors and similar trends portrayed by the forecasts resembling closely the dynamics and patterns of the actual data series. The pricing of expected spots on delivery shows an application of the model to pricing, which can be easily adopted by practitioners.

Chapter 7

Asset allocation under a multi-dimensional HMM framework

In the first part of this chapter a method is developed to optimally allocate an investment budget to two assets, where each asset price process follows a discretised geometric Brownian motion with regime-switching parameters. The filtering method and optimal parameter estimates described in chapter 3 are extended here to a multi-dimensional observation process with one underlying Markov chain as described in Elliott [49].

7.1 Filtering and parameter estimation for vector observations

Suppose $(\Omega, \mathcal{F}, P)$ denotes a probability space under which $\mathbf{x}_k$ is a homogeneous Markov chain with finite state space in discrete time ($k = 0, 1, 2...$). Again, $\mathbf{x}$ has the dynamics

$$\mathbf{x}_{k+1} = \Pi \mathbf{x}_k + \mathbf{v}_{k+1}. \tag{7.1}$$

The observation process is $d-$dimensional and the component process y^g_{k+1}, $g = 1, ..., d$, follows the dynamics

$$y^g_{k+1} = f^g(\mathbf{x}_k) + \sigma^g(\mathbf{x}_k)w^g_{k+1} \quad 1 \leq g \leq d,$$

where $\boldsymbol{f}^g = (f^g_1, f^g_2, ...f^g_N)^\top \in \mathbb{R}^N, \boldsymbol{\sigma}^g = (\sigma^g_1, \sigma^g_2, ...\sigma^g_N)^\top \in \mathbb{R}^N$ and $w^1_k, ..., w^d_k$ are $\phi(0,1)$ IID random variables, independent for each component of the observation process. Analogous to the change of probability measure technique for the one-dimensional case with equation (3.4) the real world measure P is constructed from $\bar{P}$ by defining the Radon-Nikodým derivative λ^g_l for each component g by

$$\lambda^g_l := \frac{\phi\Big[\sigma^g(\mathbf{x}_{l-1})^{-1}\big(y^g_l - f^g(\mathbf{x}_{l-1})\big)\Big]}{\sigma^g(\mathbf{x}_{l-1})\phi(y^g_l)}. \tag{7.2}$$

The Radon-Nikodým derivative of P with respect to $\bar{P}$ in the multivariate case is defined by

$$\left.\frac{dP}{d\bar{P}}\right|_{\mathcal{F}_k} = \Lambda^{dim}_k \quad \text{with} \quad \Lambda^{dim}_k := \prod_{g=1}^{d}\prod_{l=1}^{k}\lambda^g_l, \quad k \geq 1, \quad \Lambda^{dim}_0 = 1\,.$$

All components of the $d-$dimensional observation process have the same underlying Markov chain $\mathbf{x}$, all parameters therefore follow the same Markov chain and react to the same underlying information. The components are therefore correlated through the Markov chain, although the noise terms w^g are uncorrelated.

The filter equations in the multi-dimensional setting are derived in accordance with principles previously applied under the one-dimensional setting. Recall that for the Markov chain $\mathbf{x}_k$ we define $\widehat{p}^i_k := P(\mathbf{x}_k = e_i|\mathcal{F}^y_k) = \mathrm{E}\left[\langle \mathbf{x}_k, e_i\rangle\,|\mathcal{F}^y_k\right]$ and $\Xi^{dim}_k := \bar{\mathrm{E}}[\Lambda^{dim}_k \mathbf{x}_k|\mathcal{F}^y_k]$. With Bayes' theorem 2.3 we see that $\widehat{\mathbf{p}}_k = \mathrm{E}[\mathbf{x}_k|\mathcal{F}^y_k] = \frac{\bar{\mathrm{E}}[\Lambda^{dim}_k \mathbf{x}_k|\mathcal{F}^y_k]}{\bar{\mathrm{E}}[\Lambda^{dim}_k|\mathcal{F}^y_k]}$ and we have $\widehat{\mathbf{p}}_k = \frac{\Xi^{dim}_k}{\sum_{i=1}^N \langle \Xi^{dim}_k, e_i\rangle}$. The diagonal matrix $\boldsymbol{D}^{dim}$, which is utilised for the filter equations of processes from the Markov chain is defined in the multi-dimensional setting alongside equation (3.2) as

$$d^{dim}_{ij} = \begin{cases} \prod_{g=1}^{d} \frac{\phi\left(\frac{y^g_{k+1} - f^g_i}{\sigma^g_i}\right)}{\sigma^g_i \phi(y^g_{k+1})} & \text{for } i = j \\ 0 & \text{otherwise} \end{cases}.$$

The recursive relations for Ξ^{dim}_{k+1} and the filters $\eta_k(J^{(sr)}_k\mathbf{x})$, $\eta_k(O^{(r)}_k\mathbf{x})$ and $\eta_k(T^{(r)}_k\mathbf{x})$ for vector observations are given in the following theorem.

Theorem 7.1

Let $\boldsymbol{D}^{dim}$ be the diagonal matrix defined above. Then

$$\Xi_{k+1}^{dim} = \boldsymbol{\Pi} D_k^{dim} \Xi_k^{dim} \tag{7.3}$$

and

$$\begin{aligned}
\eta_l(J^{dim}{}_l^{(sr)}\mathbf{x}_l) &= \boldsymbol{\Pi}\boldsymbol{D}^{dim}(y_l)\eta_{l-1}(J^{dim}{}_{l-1}^{(sr)}\mathbf{x}_{l-1}) \\
&\quad + \langle \Xi_{l-1}, \mathbf{e}_r \rangle \prod_{g=1}^{d} \frac{\phi\left(\sigma_r^{g-1}(y_l^g - f_r^g)\right)}{\sigma_r^g \phi(y_l^g)} \pi_{sr}\mathbf{e}_s \qquad (7.4) \\
\eta_l(O^{dim}{}_l^{(r)}\mathbf{x}_l) &= \boldsymbol{\Pi}\boldsymbol{D}^{dim}(y_l)\eta_{l-1}(O^{dim}{}_{l-1}^{(r)}\mathbf{x}_{l-1}) \\
&\quad + \langle \Xi_{l-1}, \mathbf{e}_r \rangle \prod_{g=1}^{d} \frac{\phi\left(\sigma_r^{g-1}(y_l^g - f_r^g)\right)}{\sigma_r^g \phi(y_l^g)} \boldsymbol{\Pi}\mathbf{e}_r \qquad (7.5) \\
\eta_l(T^{dim}{}_l^{(r)}(g)\mathbf{x}_l) &= \boldsymbol{\Pi}\boldsymbol{D}^{dim}(y_l)\eta_{l-1}(T^{dim}{}_{l-1}^{(r)}(g)\mathbf{x}_{l-1}) \\
&\quad + \langle \Xi_{l-1}, \mathbf{e}_r \rangle \prod_{g=1}^{d} \frac{\phi(\sigma_r^{g-1}(y_l^g - f_r^g))}{\sigma_r^g \phi(y_l^g)} g(y_l^g)\boldsymbol{\Pi}\mathbf{e}_r \ . \qquad (7.6)
\end{aligned}$$

Proof

The proof is analogous to the proof of the case in the one-dimensional setting given in Appendix B.

□

With the above filter equations for the multivariate setting, the optimal parameter estimates are derived applying the EM algorithm. The analogue of theorem 3.3 for the EM estimates of the parameters in the multivariate case is the following.

Theorem 7.2

If a sequence of vector observations $y_1^g, y_2^g, \ldots, y_k^g$, $1 \le g \le d$ is available at time k and the set of parameters $\{\widehat{\pi}_{ji}, \widehat{f}_i^g, \widehat{\sigma}_i^g\}$ determines the model then the EM estimates for these parameters are given by

$$\widehat{\pi}_{ji} = \frac{\eta_k(J^{dim}{}_k^{(ji)})}{\eta_k(O^{dim}{}_k^{(i)})} \tag{7.7}$$

$$\widehat{f}_i^g = \frac{\eta_k(T^{dim}{}_k^{i}(y^g))}{\eta_k(O^{dim}{}_k^{(i)})} \tag{7.8}$$

and

$$\widehat{\sigma}_i^g = \sqrt{\frac{\eta_k(T^{dim}{}_k^{(i)}(y^g)^2) - 2f_i^g\eta_k(T^{dim}{}_k^{(i)}(y^g)) + f_i^{2}\eta_k(O^{dim}{}_k^{(i)})}{\eta_k(O^{dim}{}_k^{(i)})}}. \tag{7.9}$$

Proof

The proof is analogous to the proof of the case in the one-dimensional setting given in Appendix B.

□

Note that the filter equations and optimal parameter estimates are closely related to the one-dimensional setting, since one underlying Markov chain governs the parameters. The change of measure has to be adapted to the multi-dimensional setting by defining Λ_k^{dim} as the product of $\Lambda_k's$ for each component of the observation process. The other derivations of filters and parameter estimates in the multi-dimensional setting follow straightforward.

7.2 Forecast of indices

The filtering and parameter estimation approach for vector observations is now utilised to develop an optimal strategy for an asset allocation problem. Suppose a fund manager has to decide whether to invest in growth or value funds. A 'value-driven' investment fund tends to include investments in large, stable companies, which tend to have a steady return. A 'growth-driven' investment fund on the other hand tends to be more risky and includes investments in companies which might have lower returns today but significantly better prospects for the future. The optimal strategy, which is developed here will be based on return forecasts of two popular stock indices, the NASDAQ and Dow Jones. These indices can be seen as typical performances of growth stocks, which are mirrored by the NASDAQ index, and value stocks, mirrored by the Dow Jones index.

The two indices are treated here as a two-dimensional observation process. The filtering method together with the resulting optimal parameter estimation from section 7.1 is implemented on NASDAQ and Dow Jones weekly data for a period of about ten years, from March 1997 until June 2007. The vector process of returns for both indices $y^{Ndq} := \ln \frac{Nasdaq(k+1)}{Nasdaq(k)}$ and $y^{DJ} := \ln \frac{DowJ(k+1)}{DowJ(k)}$ considered here is

therefore given by

$$y_{k+1}^{Ndq} = f^{Ndq}(\mathbf{x}_k) + \sigma^{Ndq}(\mathbf{x}_k) w_{k+1}^{Ndq} \tag{7.10}$$

and

$$y_{k+1}^{DJ} = f^{DJ}(\mathbf{x}_k) + \sigma^{DJ}(\mathbf{x}_k) w_{k+1}^{DJ} \tag{7.11}$$

where $\boldsymbol{f}^{Ndq} = (f_1^{Ndq}, f_2^{Ndq}, ...f_N^{Ndq})^\top \in \mathbb{R}^N, \boldsymbol{\sigma}^{Ndq} = (\sigma_1^{Ndq}, \sigma_2^{Ndq}, ...\sigma_N^{Ndq})^\top \in \mathbb{R}^N$, $\boldsymbol{f}^{DJ} = (f_1^{DJ}, f_2^{DJ}, ...f_N^{DJ})^\top \in \mathbb{R}^N, \boldsymbol{\sigma}^{DJ} = (\sigma_1^{DJ}, \sigma_2^{DJ}, ...\sigma_N^{DJ})^\top \in \mathbb{R}^N$, and w_k^{Ndq} and w_k^{DJ} are $\phi(0,1)$ IID random variables, independent of each other. The indices at time k are given by $Nasdaq(k)$ and $DowJ(k)$, respectively. In both processes, the parameters are governed by the same Markov chain $\mathbf{x}_k$ in discrete time. The underlying Markov chain is able to switch between economic regimes, therefore the parameters of both observation processes driven by this Markov chain are regime-switching. In the implementation of filtering and parameter estimation the number of states N is set to $N = 3$. This ensures that the model dynamics are flexible enough to switch between different economic regimes within a reasonably low number of parameters. The one-step ahead forecast of both indices is given analogous to the one-dimensional case in equation (3.18) by

$$\mathrm{E}\left[Nasdaq(k+1)|\mathcal{F}_k^y\right] = Nasdaq(k) \sum_{i=1}^{N} \langle \widehat{\mathbf{x}}_k, e_i \rangle \exp\left(f_i^{Ndq} + \frac{\sigma_i^{Ndq^2}}{2}\right) \tag{7.12}$$

and

$$\mathrm{E}\left[DowJ(k+1)|\mathcal{F}_k^y\right] = DowJ(k) \sum_{i=1}^{N} \langle \widehat{\mathbf{x}}_k, e_i \rangle \exp\left(f_i^{DJ} + \frac{\sigma_i^{DJ^2}}{2}\right). \tag{7.13}$$

Figure 7.1 depicts the one-step ahead forecasts of both indices within the considered time period. In the implementation the parameters of both processes are updated every ten data steps, that means that the parameters of the observation processes with weekly data are updated roughly after two and a half months. Both resulting forecasts are very close to the actual data. The evolution of parameters is shown in Figure 7.2. All four parameters in the indices' models are governed by the same

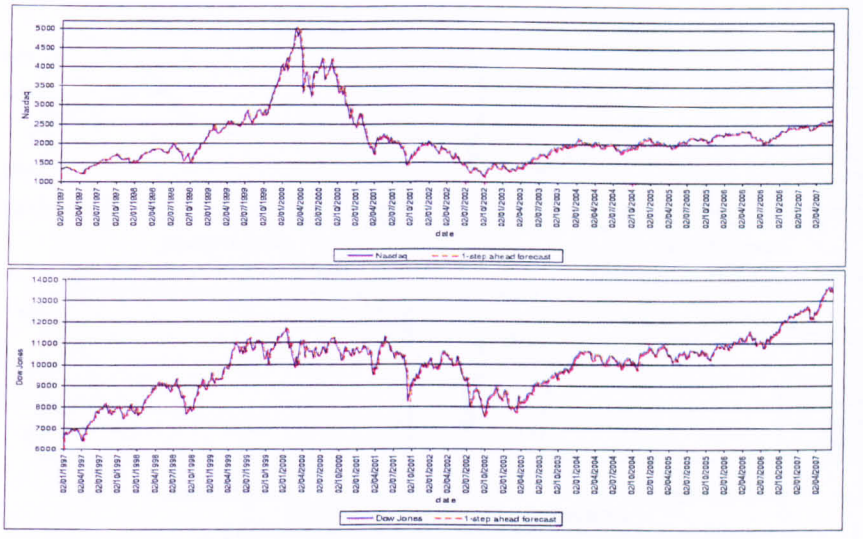


Figure 7.1: Actual data and one-step ahead forecasts for the NASDAQ and Dow Jones indices

Markov chain, whose transition probabilities are depicted in the lower part of Figure 7.2.

These index levels forecasts are utilised as signals to find an optimal investment strategy: how much to invest in NASDAQ and the remaining wealth to invest in Dow Jones. We assume no transaction costs and no consumption for simplicity. Consider the forecasted returns for the next time period for each index. The forecasted returns are divided by the realised volatility of each index from the previous 20 days, and these risk-adjusted forecasted returns then provide a signal for the investment decision in the next period. In each time step the signals of both indices for the next time step are compared. The investor invests in the index which has the higher forecasted risk-adjusted return for the next time step. This resulting investment strategy is compared with the two pure investment strategies of either purely investing in NASDAQ or Dow Jones. The performance of the switching strategy induced by forecasted return signals is compared with the pure investment strategy for weekly data beginning 1997 until 2007. The time series data is divided into 41 intervals, each corresponds to one yearly quarter. The optimal switching strategy is calculated for each of these quarters. An initial investment budget of \$100 is assumed. The investment strategies and the comparison to the

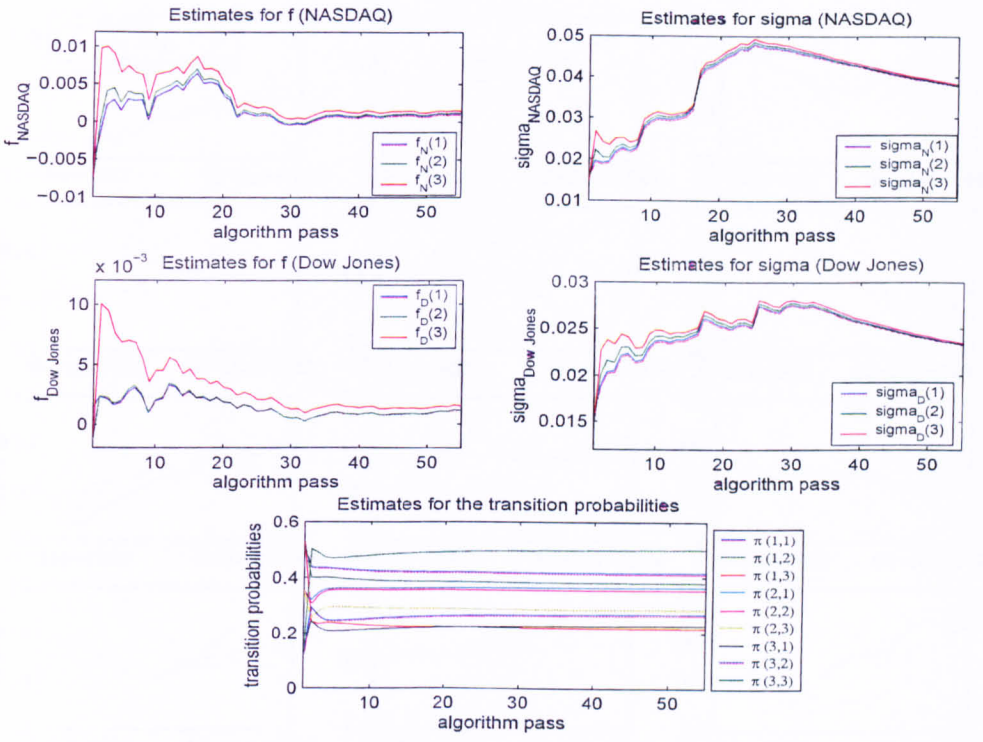


Figure 7.2: Evolution of parameters for a two-dimensional observation process

pure strategies can be seen in Figures 7.3 and 7.4 below. A better performance of the switching strategy compared to both 'pure' index strategies in the sense of a higher return at the end of the considered quarters can be seen in 12 quarters (about 27% of considered data sets), a higher return than in at least one of the 'pure' strategies is achieved in about 76 % of the considered quarters.

The overall performance in each quarter of the investment strategy is compared to the overall performance of each index through the functions

$$\begin{aligned} X^{Ndq}_{quart} &:= \ln \frac{SwS_{quart}}{100} - \ln \frac{NdqS_{quart}}{100} \\ X^{DJ}_{quart} &:= \ln \frac{SwS_{quart}}{100} - \ln \frac{DJS_{quart}}{100} \end{aligned},$$

where $quart = 1, ..., 41$ is the number of the quarter, SwS_{quart} denotes the value of the investment through the switching strategy at the end of quarter $quart$ and $NdqS_{quart}$ and DJS_{quart} denote the value of the investment through the pure NASDAQ and Dow Jones strategy at the end of the quarter $quart$, respectively. The

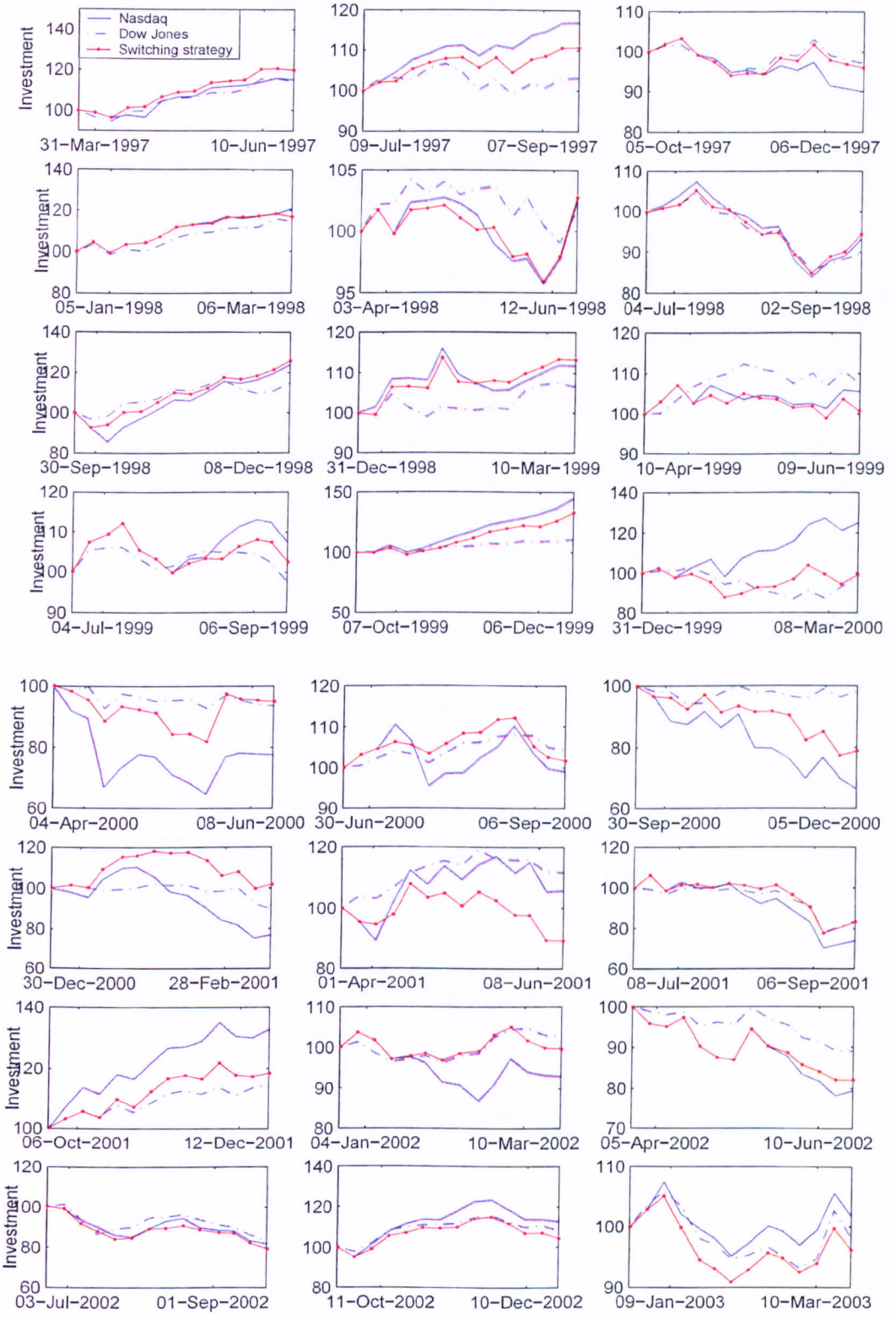


Figure 7.3: Comparison of pure and switching investment strategies for 24 quarters from 1997 to 2003

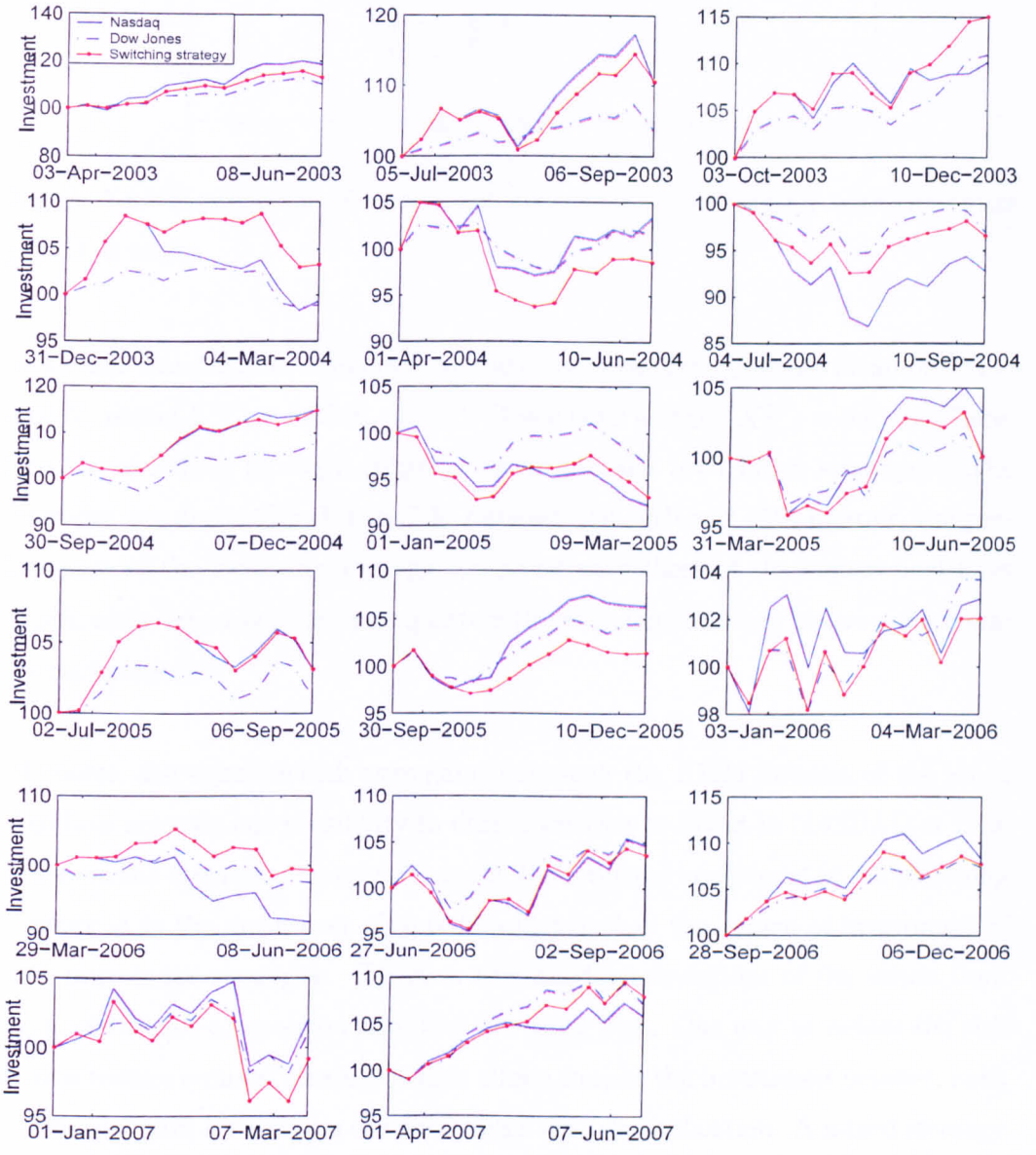


Figure 7.4: Comparison of pure and switching investment strategies for 17 quarters from 2003 to 2007

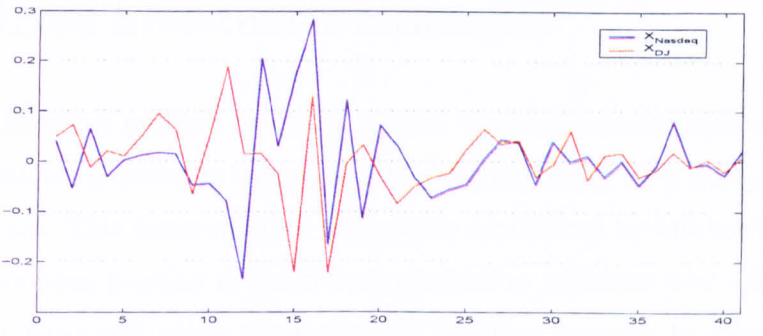


Figure 7.5: Comparison of investment performance per quarter over the entire period of study

mean and standard deviation over all quarters of the functions are $\text{mean}(X^{Ndq}) = 0.0037$, $\text{mean}(X^{DJ}) = 0.0048$, $\text{std}(X^{Ndq}) = 0.0892$ and $\text{std}(X^{DJ}) = 0.0725$. Therefore the switching strategy outperforms on average both index strategies. The functions are depicted in Figure 7.5. Between the 10th and 15th quarter, the performance of the switching strategy compared with those of the indices fluctuates significantly, whereas in the last quarters the switching strategy shows a more stable performance.

Of course, the signals which were gained through the HMM forecast of the levels of indices are only one possibility to decide whether to invest in NASDAQ or Dow Jones. In the absence of transaction costs, the return of investment of the switching strategy is in the mean over all quarters higher than the return of investment of both pure index strategies. However, the standard deviations of the values from the performance comparison functions are quite high. One way to reduce the risk is to introduce a mixed strategy, which allows to split the investment between both indices and gain a reduction of risk through that diversification. A mixed strategy for optimal investments which entails the derivation of optimal allocation weights between two assets is presented in the next section.

7.3 Mixed investment strategies

The asset allocation problem can be seen as a trade-off between risk and return. The more risk the investor is willing to take the more likely he is going to receive a higher return. This mean-variance theory was established by Markowitz [112] in 1952 and has been popular in portfolio optimisation problems ever since. Elliot and van der Hoek [58] use a mean-variance function in order to solve an asset allocation problem within an HMM setting, where the returns include observable characteristics. Luo [106] extends their approach to a multi-time framework. Both approaches consider unobservable and observable factors and derive optimal weights in their models. Another HMM for an optimal portfolio choice was developed by Elliott and Hinz [52]. In their work, price changes of portfolio assets are modelled by an HMM and optimal trading times are determined. We formulate this mean-variance problem within this HMM through a function MV, which is a linear combination of return and risk. Suppose y^{Ndq} and y^{DJ} are the NASDAQ and Dow Jones returns, respectively. The function we wish to maximise with respect to the weights $\boldsymbol{w} = (w_1, w_2)$ is defined by

$$\begin{aligned} MV(\boldsymbol{w}) &= \nu \,\mathrm{E}\big[w_1 y_{k+1}^{Ndq} + w_2 y_{k+1}^{DJ} \mid \mathcal{F}_k^y\big] \\ &\quad -\mathrm{Var}\big[w_1 y_{k+1}^{Ndq} + w_2 y_{k+1}^{DJ} \mid \mathcal{F}_k^y\big] \ . \end{aligned} \tag{7.14}$$

The following theorem gives the optimal weights in the context of the present portfolio optimisation problem considered in this chapter.

Theorem 7.3

Let $\nu > 0$ be the risk aversion factor. The optimal weight w_1 is given by

$$w_1 = \frac{\nu(\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle) + 2\langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2}{2(\langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2 + \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2)}, \tag{7.15}$$

and the optimal weight w_2 is given by $w_2 = 1 - w_1$. The risk aversion factor ν has to fulfill

$$\nu \leq \frac{2\langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2}{\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle}$$

for $\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle \ > \ 0$

and

$$\nu \leq -\frac{2\langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2}{\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle}$$

$$\text{for } \langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle \;<\; 0.$$

Proof

First, let us consider the mean-variance function MV

$$\begin{aligned} MV(\boldsymbol{w}) &= \nu E\big[w_1 y_{k+1}^{Ndq} + w_2 y_{k+1}^{DJ} \mid \mathcal{F}_k^y\big] - \mathrm{Var}\big[w_1 y_{k+1}^{Ndq} + w_2 y_{k+1}^{DJ} \mid \mathcal{F}_k^y\big] \\ &= \nu E\big[w_1 y_{k+1}^{Ndq} + w_2 y_{k+1}^{DJ} \mid \mathcal{F}_k^y\big] - E\big[(w_1 y_{k+1}^{Ndq} + w_2 y_{k+1}^{DJ})^2 \mid \mathcal{F}_k^y\big] \\ &\quad + \Big[E\big[w_1 y_{k+1}^{Ndq} + w_2 y_{k+1}^{DJ} \mid \mathcal{F}_k^y\big]\Big]^2 \\ &= \nu w_1 \hat{y}_{k+1}^{Ndq} + \nu w_2 \hat{y}_{k+1}^{DJ} - E\big[(w_1 y_{k+1}^{Ndq})^2 + 2 w_1 y_{k+1}^{Ndq} w_2 y_{k+1}^{DJ} + (w_2 y_{k+1}^{DJ})^2 \mid \mathcal{F}_k^y\big] \\ &\quad + (w_1 \hat{y}_{k+1}^{Ndq})^2 + 2 w_1 \hat{y}_{k+1}^{Ndq} w_2 \hat{y}_{k+1}^{DJ} + (w_2 \hat{y}_{k+1}^{DJ})^2 \end{aligned}$$

where $\hat{y}_{k+1} = E\big[y_{k+1} \mid \mathcal{F}_k^y\big]$. *Utilising the fact that* $\hat{y}_{k+1}^{Ndq} = \langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle$ *and* $\hat{y}_{k+1}^{DJ} = \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle$ *we have*

$$\begin{aligned} MV(\boldsymbol{w}) &= \nu w_1 \langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle + \nu w_2 \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle \\ &\quad - E\big[(w_1 y_{k+1}^{Ndq})^2 \mid \mathcal{F}_k^y\big] - E\big[2 w_1 y_{k+1}^{Ndq} w_2 y_{k+1}^{DJ} \mid \mathcal{F}_k^y\big] - E\big[(w_2 y_{k+1}^{DJ})^2 \mid \mathcal{F}_k^y\big] \\ &\quad + (w_1 \langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle)^2 + 2 w_1 \langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle w_2 \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle \\ &\quad + (w_2 \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle)^2 . \end{aligned}$$

With

$$E\big[(w_1 y_{k+1}^{Ndq})^2 \mid \mathcal{F}_k^y\big] = w_1^2 (\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2 + \langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2) \tag{7.16}$$

and

$$E\big[(w_2 y_{k+1}^{DJ})^2 \mid \mathcal{F}_k^y\big] = w_2^2 (\langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2 + \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2) \tag{7.17}$$

we get

$$\begin{aligned} MV(\boldsymbol{w}) &= \nu w_1 \langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle + \nu w_2 \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle \\ &\quad - w_1^2 \langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2 - w_2^2 \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2 . \end{aligned} \tag{7.18}$$

Now, substituting w_2 with $1 - w_1$ leads to

$$\begin{aligned} MV(w_1) &= \nu w_1(\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle) + \nu \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle \\ &\quad -w_1^2(\langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2 + \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2) \\ &\quad -\langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2 + 2w_1 \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2 . \end{aligned} \tag{7.19}$$

To maximise $MV(w_1)$ subject to the constraints $w_1 \leq 1$ and $-w_1 \leq 0$ we have the Lagrangian $L(w_1, v_1, v_2)$

$$\begin{aligned} L &= \nu w_1(\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle) + \nu \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle \\ &\quad -w_1^2(\langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2 + \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2) - \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2 \\ &\quad +2w_1 \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2 + v_1(1 - w_1) + v_2 w_1 , \end{aligned} \tag{7.20}$$

which gives the optimality conditions

$$\begin{aligned} &\nu(\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle) - 2w_1(\langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2 + \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2) \\ &\quad +2\langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2 - v_1 + v_2 = 0 \\ &\quad v_1(1 - w_1) = 0 \\ &\qquad v_2 w_1 = 0 \\ &\qquad w_1 \leq 1 \\ &\qquad -w_1 \leq 0 \\ &\qquad v_1, v_2 \geq 0 . \end{aligned}$$

These are fulfilled by the optimal weight w_1

$$w_1 = \frac{\nu(\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle) + 2\langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2}{2(\langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2 + \langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2)} , \tag{7.21}$$

with

$$-\frac{2\langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2}{\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle} \leq \nu \leq \frac{2\langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2}{\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle}$$

for $\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle \;>\; 0$

and

$$\frac{2\langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2}{\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle} \leq \nu \leq -\frac{2\langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2}{\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle}$$

$$\text{for } \langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle \ < \ 0.$$

In the case where $w_1 = 0$ *we further have*

$$\nu \leq -\frac{2\langle \boldsymbol{\sigma}^{DJ}, \widehat{\mathbf{x}}_k \rangle^2}{\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle}$$

and for the case $w_1 = 1$

$$\nu \geq \frac{2\langle \boldsymbol{\sigma}^{Ndq}, \widehat{\mathbf{x}}_k \rangle^2}{\langle \boldsymbol{f}^{Ndq}, \widehat{\mathbf{x}}_k \rangle - \langle \boldsymbol{f}^{DJ}, \widehat{\mathbf{x}}_k \rangle}.$$

□

This asset allocation strategy, where the optimal weights depend on the optimal parameters and the state of the Markov chain estimated in the HMM framework, is now implemented on the NASDAQ and Dow Jones data sets. Again, each data set is divided into the previously defined 41 quarters. For each quarter, the optimal weights in each time step for NASDAQ and Dow Jones are obtained. The risk aversion factor ν is set to 0.0697 to ensure that the constraints are fulfilled for each algorithm pass. The optimal calculated weights which are used for the optimal investment strategies can be seen in Figure 7.6. They are derived from the estimated parameters in the HMM setting for both indices. The evolution of the investments in the 41 quarters is depicted below in Figures 7.7 and 7.8.

The Figures show clearly that, in terms of the mean returns, the mixed strategy with the optimal calculated weights does not necessarily outperform the pure investment and switching strategies. However, since we introduced a mean-variance utility function, which maximises the return by minimising the variance, the mean return and variance of each strategy for the quarters is compared in Figure 7.9. Here it can be seen, that the mean-variance strategy reduces the risk of the investment compared to the pure strategies.

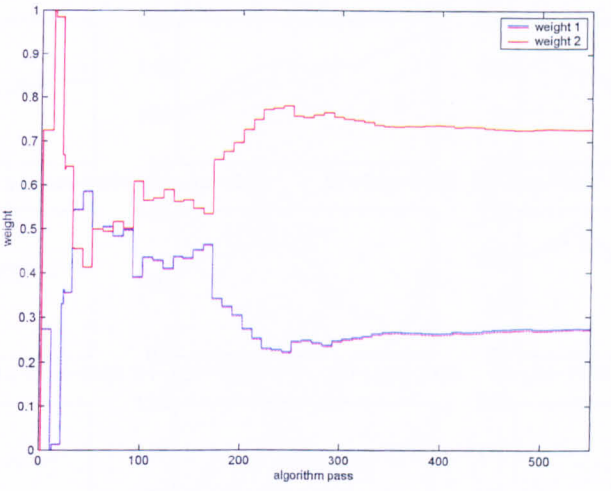


Figure 7.6: Optimal weights implied by the NASDAQ and Dow Jones data for the mixed strategy under the HMM setting

7.4 HMM based scenario generation for an asset allocation problem

Stochastic programming is an important technique that can be implied to calculate the optimal weights for investments in a portfolio optimisation problem. The decision maker has a choice in which assets to invest his wealth in order to achieve an optimal combination of risk and return. Since the outcome of the investment is uncertain, techniques are developed to generate scenarios with numerous possible future returns, which aim to provide a realistic outlook into future performances. The problem formulation for stochastic optimisation problems relies on different scenarios which need to be generated so the problem can be efficiently solved by stochastic programming algorithms. This is an area of active research and an overview of financial optimisation problems can be found for example in Mulvey [119].

Various models for optimal portfolio selection have been proposed, starting with the popular mean-variance approach by Markowitz [112]. More recently portfolio optimisation problems include more sophisticated risk measures. A popular risk measure is Value-at-Risk (VaR), which indicates the maximum amount one can

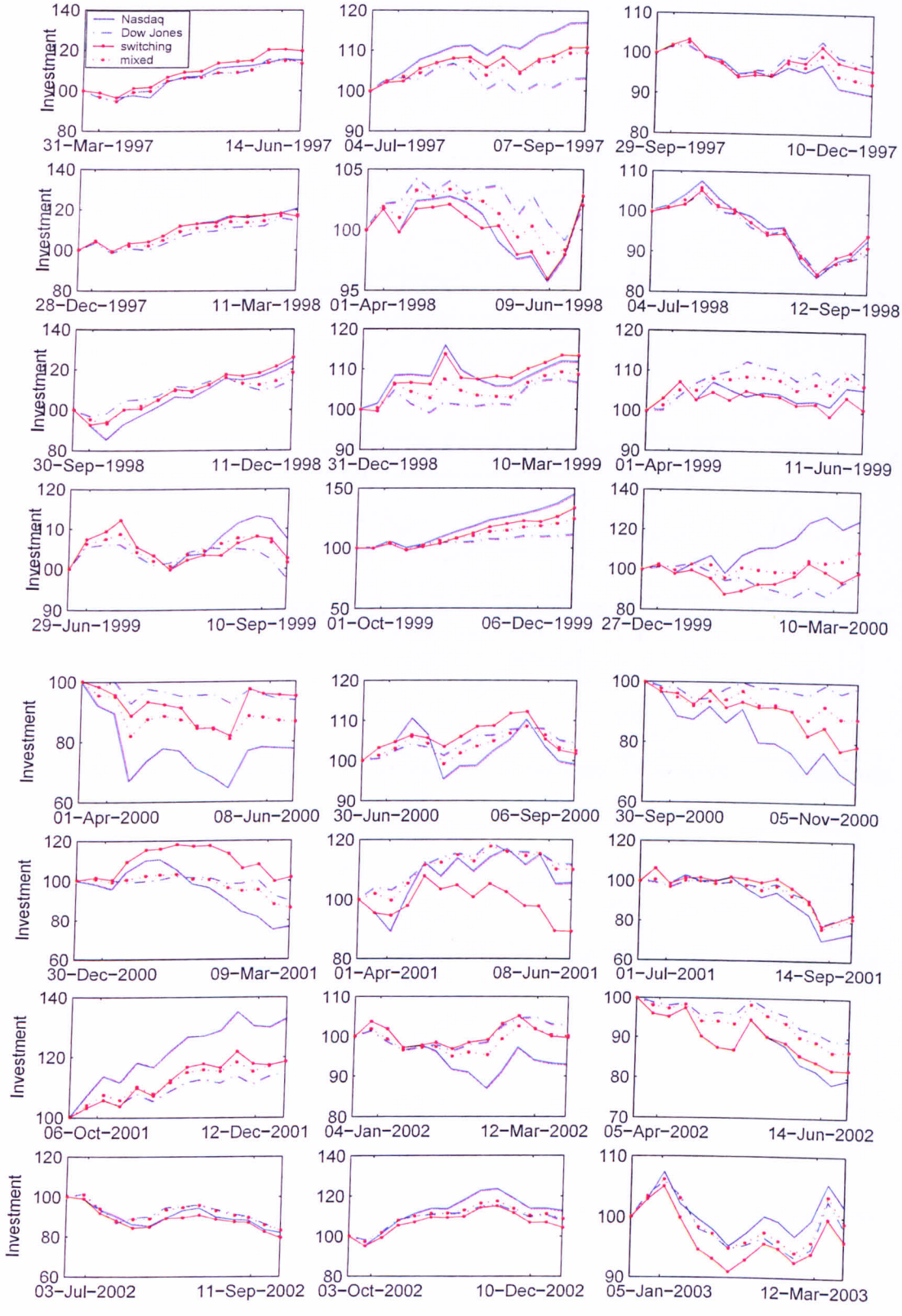


Figure 7.7: Performance results of pure, switching and mixed investment strategies for 24 quarters from 1997 to 2003

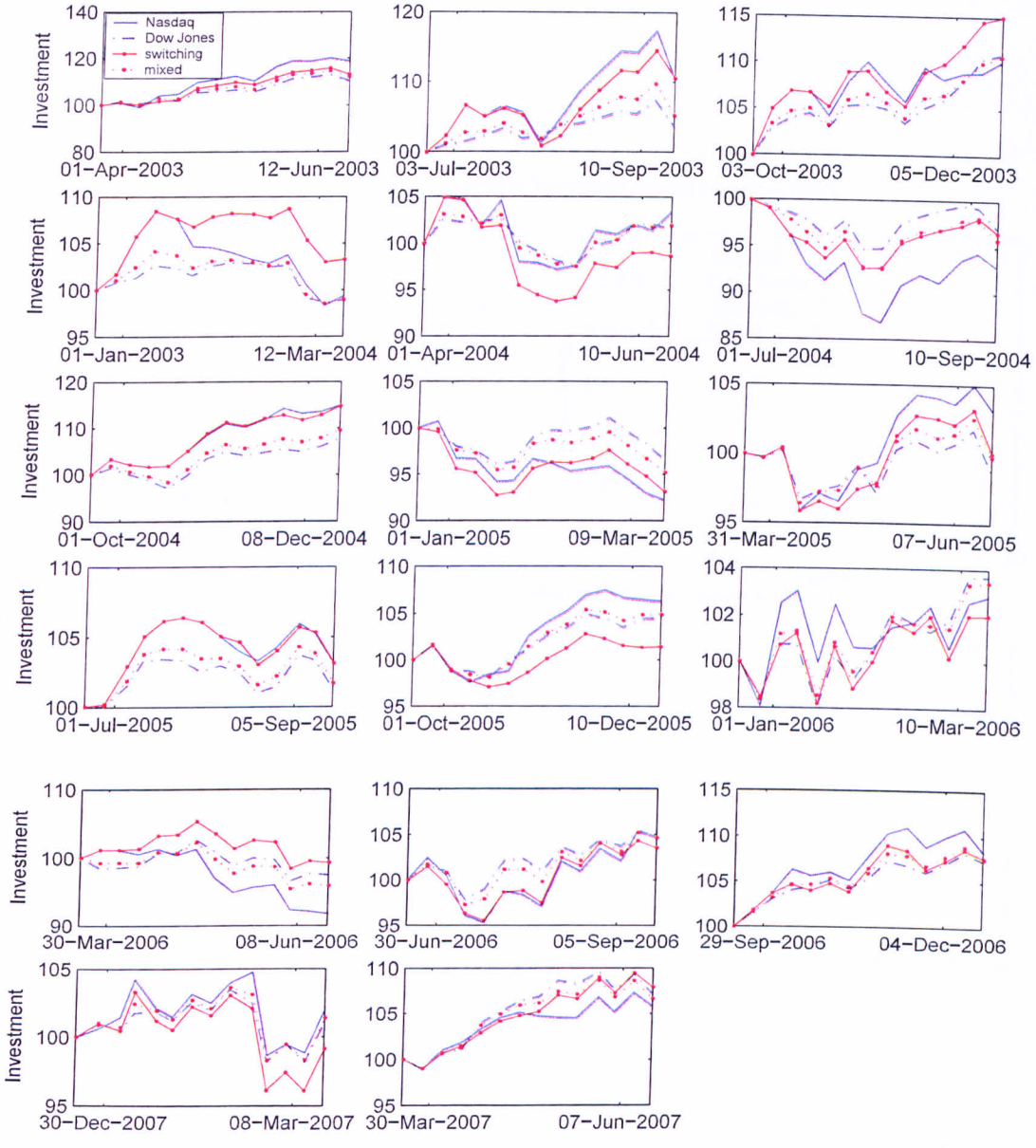


Figure 7.8: Performance results of pure, switching and mixed investment strategies for 17 quarters from 2003 to 2007

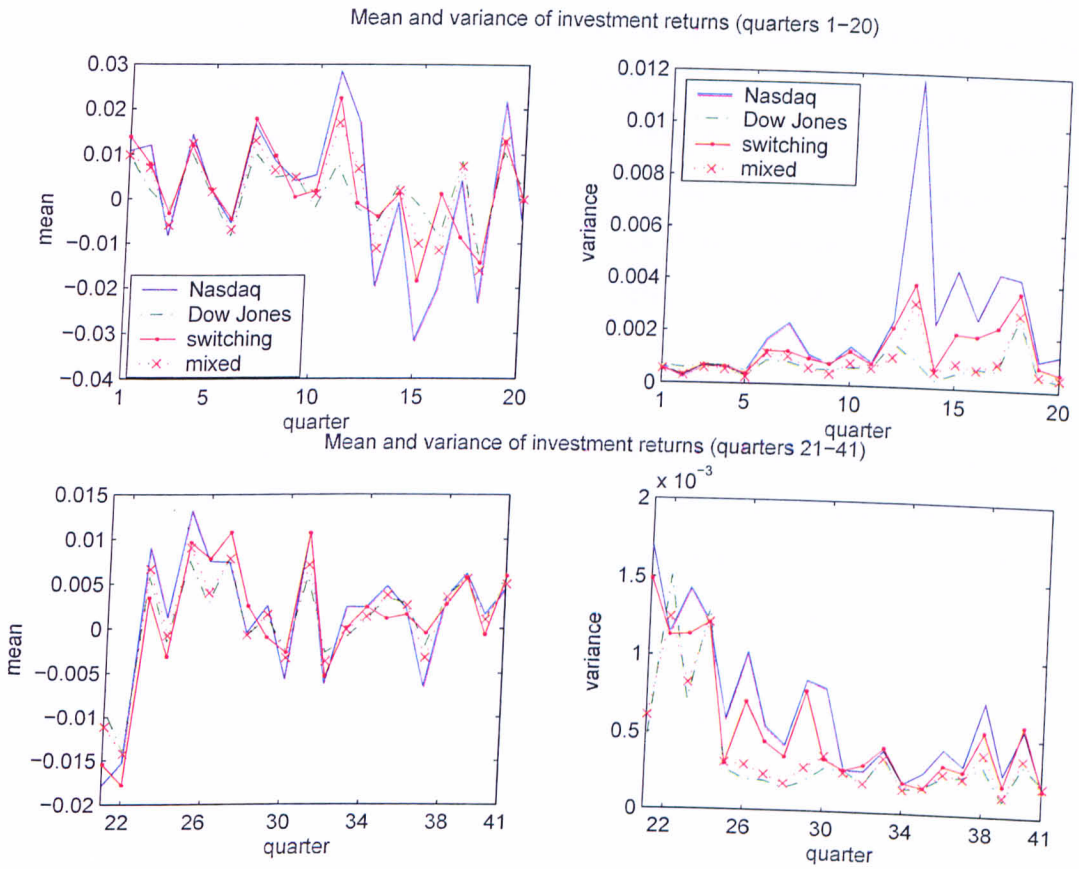


Figure 7.9: Comparison of the mean and variance of the returns in each quarter for the pure index, switching and mixed investment strategies

loose at a particular confidence level. As pointed out by Artzner et al. [8], one drawback of this measure in the application to stochastic optimisation problems is that it is non-smooth and non-convex and can therefore lead to multiple local extrema. On the other hand, the conditional Value-at-Risk (CVaR), also known as Mean Excess Loss or Mean shortfall, is defined as the conditional expectation of portfolio returns below the VaR return and is a coherent risk measure (see Pflug [127]). Rockafellar and Uryasev [130] as well as Krokhmal et al. [96] developed CVaR models for portfolio optimisation, where on the one hand one focuses on minimising risk with a given minimum level of portfolio return, whilst on the other hand one considers maximising return with a given maximum level of risk. Krokhmal et al. [96] showed, that both formulations lead to the same efficient frontier of the portfolio under the CVaR framework. Both model formulations are applied to the decision problem here. So, the investor can therefore choose if he wants to fix the risk or return level in the optimisation.

One important issue for various portfolio optimisation problems is the scenario generation for the underlying variables. Different approaches have been used to generate scenarios in CVaR decision models. Andersson et al. [3] used Monte Carlo simulation for a credit risk optimisation problem, Topaloglou et al. [144] generated scenarios with a principal component analysis for asset allocation in a CVaR framework. Here, a scenario generator is developed which is based on an HMM. The underlying hidden information for asset prices and indices can symbolise different stages of a business cycle, namely, expansion, peak, recession, trough and recovery, which influence the price movements. The scenarios are generated based on optimal parameter estimates within the HMM, both in a one-dimensional and multivariate setting as described in chapter 3 and section 7.1. Once optimal parameters are estimated using historical information of the prices, scenario paths are generated using parameter estimates and the filtered transition probability of the underlying Markov chain.

In the next section, the investment problem is described and the optimisation problem is formulated. The scenario generation technique is derived in section 7.5 for independent as well as vector observation processes. A step-by-step description

of the scenario generator is given which can be applied to a variety of financial time series. Section 7.6 shows the results of the decision problem solved with different sets of scenarios and an extension to other problem formulations. The main finding shows that the scenarios generated through the HMM method are stable for various decision problem formulations. Optimal solutions are obtained for both CVaR models, one with given minimal return and the other with given maximal risk.

7.5 Investment problem in a CVaR framework

The asset allocation problem considered here is to find the optimal ratios for investing in gold spots, FTSE 100 stocks or a risk-free bank account. An investment in gold is traditionally seen as a long-term investment that can play a role in hedging against inflation and political or economic problems. Like other commodities, gold is traded on spot and future markets, but it has specific characteristics which are not common within commodity markets. Due to its role as a global currency, prices on spot markets are global. Gold supply and demand do not depend on seasonality, it has a low risk of supply interruption and low storage and insurance costs. Furthermore gold has no risk of spoilage and the consumption level relative to inventory is low. Unlike price processes for other commodities, a convenience yield is not included in the price process due to the distinctive gold features. The performances of the gold spot prices and the FTSE 100 index are estimated under different scenarios, which are generated with parameters estimated by the HMM filtering method.

The decision in this modelling framework is based on calculating the conditional value at risk (CVaR) for the portfolio (see [132] for more details). For a given minimal portfolio return the portfolio with the lowest CVaR is chosen by the optimisation algorithm. We follow the definition by Krokhmal et al. [96]. Let xp be a decision vector that can be interpreted as a portfolio from the set of available portfolios XP. Suppose $\mathrm{ls}(xp, y)$ is the loss associated with the decision vector xp, where y are the uncertainties that can affect the loss. The probability of $\mathrm{ls}(xp, y)$

not exceeding a threshold tr is given by

$$\Psi(xp, b) = \int_{\mathrm{ls}(xp,y)\leq tr} p(y)dy,$$

where $p(y)$ denotes the density of y. The $\alpha-$VaR and $\alpha-$CVaR for any specified probability level α in $(0, 1)$ is denoted by $b_\alpha(xp)$ and $c_\alpha(xp)$, respectively and defined as

$$b_\alpha(xp) = \min\{tr \in \mathbb{R} : \Psi(xp, tr) \geq \alpha\} \tag{7.22}$$

$$c_\alpha(xp) = (1-\alpha)^{-1} \int_{\mathrm{ls}(xp,y)\geq b_\alpha(xp)} \mathrm{ls}(xp, y)p(y)dy \tag{7.23}$$

where $c_\alpha(xp)$ is the conditional expectation of the loss associated with the portfolio relative to that loss being $b_\alpha(x)$ or greater. To characterise $b_\alpha(xp)$ and $c_\alpha(xp)$ let us define a function $F_\alpha(xp, tr) : XP \times \mathbb{R} \mapsto \mathbb{R}$. In particular,

$$F_\alpha(xp, tr) = tr + (1-\alpha)^{-1} \int_{y\in\mathbb{R}^n} \left[\mathrm{ls}(xp, y) - tr\right]^+ p(y)d(y) \tag{7.24}$$

with $[t]^+ = \max\{t, 0\}$. As a function of tr, $F_\alpha(xp, tr)$ is convex and continuously differentiable. The $\alpha-$CVaR can then be determined from

$$c_\alpha(xp) = \min_{tr\in\mathbb{R}} F_\alpha(xp, tr). \tag{7.25}$$

Minimising the $\alpha-$CVaR is equivalent to minimising $F_\alpha(xp, tr)$ over all $(xp, tr) \in XP \times \mathbb{R}$ (see Krokhmal et al. [96], theorem 2). So,

$$\min_{xp\in XP} c_\alpha(xp) = \min_{(xp,tr)\in X\times\mathbb{R}} F_\alpha(xp, tr). \tag{7.26}$$

The discretised version of the problem leads to the approximation

$$\tilde{F}_\alpha(xp, tr) = tr + (1-\alpha)^{-1} \sum_{\mathrm{scen}=1}^{SCEN} Prob_{\mathrm{scen}}[\mathrm{ls}(xp, y_{\mathrm{scen}}) - tr]^+ \tag{7.27}$$

where $Prob_{\mathrm{scen}}$ are the probabilities of scenario y_{scen}.

The defining elements of the optimisation problem are:

- **Indices**

  SCEN : scenario set

- **Data**

  $no\_scenarios$: number of scenarios generated for spot and stock returns

  $ReturnSpot_{\text{scen}}$: gold spot price return for scenario scen

  $ReturnFTSE_{\text{scen}}$: FTSE100 return for scenario scen

  $Prob_{\text{scen}} := \frac{1}{no\_scenarios}$ probability of scenario scen

  $ReturnBond$: Return on bank account

  $MReturnSpot$: Mean gold price return over all scenarios scen

  $MReturnFTSE$: Mean FTSE100 return over all scenarios scen

  β_{CV} : Confidence level for CVaR

  $MinPortRet$: minimal required level of expected portfolio return

- **Decisions**

  $BuySpots$: percentage of budget invested in gold spots

  $BuyFTSE$: percentage of budget invested in FTSE100 stocks

  $BuyBond$: percentage of budget invested in bank account

  $negdev_{\text{scen}}$: difference between VaR and portfolio return in scenario scen

  α_V : level of VaR

  Z : objective function

- **Objective:** Minimise CVaR

$$Z = \alpha_V + \frac{1}{\beta_{CV}} * \sum_{\text{scen} \in SCEN} Prob_{\text{scen}} * negdev_{\text{scen}}$$

- **subject to the following constraints**

  1. Weight constraint:

     BuySpots + BuyFTSE + BuyBond = 1

  2. Minimal portfolio return:

     BuySpots * MReturnSpots + BuyFTSE * MReturnFTSE + BuyBond * ReturnBond $\geq MinPortRet$

3. CVaR constraint:

   $\text{BuySpots} * \text{ReturnSpot}_{\text{scen}} + \text{BuyFTSE} * \text{ReturnFTSE}_{\text{scen}}$

   $+\text{BuyBond} * \text{ReturnBond} + \alpha_V + negdev_{\text{scen}} \geq 0 \ \forall \ \text{scen} \in \text{SCEN}$

4. Positivity constraints:

   $negdev \geq 0$

   $\text{BuySpots} \geq 0, \ \text{BuyFTSE} \geq 0, \ \text{BuyBond} \geq 0$

The FTSE 100 index, similar to stock prices, can be modelled by a geometric Brownian motion. Since gold spot prices unlike other commodity prices do not show seasonality etc., a convenience yield is not included in the price process and it can therefore be modelled by a geometric Brownian motion. The parameter estimation and forecast calculated here are based on an observation process whose dynamics follow a discretised geometric Brownian motion. Thus, the scenarios are generated under the HMM framework. The parameters in the observation processes are governed by the underlying Markov chain, which is not directly observable. For the scenario generation, the log returns of both price processes are assumed to follow the dynamics in discrete time as given for the gold spot price in chapter 3, i.e.

$$y_{k+1} = f(\mathbf{x}_k) + \sigma(\mathbf{x}_k) w_{k+1}. \tag{7.28}$$

The parameters f and σ are governed by the Markov chain $\mathbf{x}$ in discrete time and are therefore able to switch between different regimes. The $w_k's$ are a sequence of IID standard normal random variables independent of $\mathbf{x}$.

Under the real world probability measure P, the Markov chain $\mathbf{x}$ again has the dynamics (3.3). We adopt filters for the Markov chain and related quantities developed in section 3.2 and recursive optimal parameter estimates for these one-dimensional regime-switching processes derived in section 3.3. Recall that the recursive optimal parameter estimates for the transition probabilities π_{ji}, the mean f_i and the variance σ_i of the observation process are

$$\widehat{\pi}_{ji} = \frac{\widehat{J}_k^{(ji)}}{\widehat{O}_k^{(i)}}, \tag{7.29}$$

$$\widehat{f}_i = \frac{\widehat{T}_k^{(i)}}{\widehat{O}_k^{(i)}} \tag{7.30}$$

and

$$\widehat{\sigma}_i = \sqrt{\frac{\widehat{T}_k^{(i)}(y^2) - 2\widehat{f}_i\widehat{T}_k^{(i)}(y) + \widehat{f}_i^2\widehat{O}_k^{(i)}}{\widehat{O}_k^{(i)}}}. \tag{7.31}$$

With these optimal parameter estimates one-step ahead forecasts for the data series can be computed. A three-state Markov chain is considered for the scenario generation. For most actual data series three states are sufficient to capture different states of the economy without overfitting the model. The input for the scenario generator are historical time series data and a period of one year is sufficient to estimate optimal parameters for the next time step.

The scenario generation takes the following steps:

1. Estimate optimal parameters for the time series with an underlying three-state Markov chain.

2. Estimate the optimal transition probability matrix for the considered time series.

3. Draw normally distributed random variables, which are used for generating the white noise part.

4. Generate scenarios for the next time step:
   a) the Markov chain for the next time step is calculated with its expectation conditioned on the filtration $\mathcal{F}_k^y$. We therefore have $E[\mathbf{x}_{k+1}|\mathcal{F}_k^y] = \Pi\hat{\mathbf{x}}_k$ with $\hat{\mathbf{x}}_k = E[\mathbf{x}_k|\mathcal{F}_k^y]$.

   b) the parameter values for the next time step are calculated as the scalar product between the expected Markov chain $\mathbf{x}_{k+1}$ and the estimated optimal parameters,

   $$f_{scen} = \langle \boldsymbol{f}, \mathbf{x}_{k+1} \rangle \text{ and } \sigma_{scen} = \langle \boldsymbol{\sigma}, \mathbf{x}_{k+1} \rangle.$$

   c) the scenarios for the prices in the next time step are created with the discretised version of the geometric Brownian motion using the parameters stated above

   $$S_{scen} = C(k) * \exp(f_{scen} + \sigma_{scen} * w_{scen}),$$

where $C(k)$ denotes the last actual data point of the observation process and w denotes the IID standard normally distributed random variable independent of the Markov chain for Scenario *scen*.

d) the required log returns of the data series are then calculated for each scenario using the estimated prices calculated in the different scenarios above.

In the following sections, two different HMM frameworks for the scenario generations are considered. In the first setting both time series are assumed to be independent. The optimal parameters for each process are derived in separate algorithms. The second setting assumes a dependency between the time series. Both observation processes are governed by the same Markov chain and are therefore estimated as vector observations. The parameter estimation in this case follows the theory in section 7.1, formulae (7.7), (7.8) and (7.9).

7.5.1 Scenarios under an independent observation process setting

The underlying historical data series for the optimal parameter estimation are time series for daily prices of gold spots and the FTSE 100 index over a one-year period. The algorithm for the optimal parameter estimates runs 24 times on these data sets. First the time series are assumed to be independent; therefore, both data sets follow independent Markov chains. For each data set the parameters are updated when new information arrives after batches of 10 data points. With this self-tuning algorithm for the regime-switching observation process optimal parameter estimates are calculated and the transition probability matrix of the Markov chain $\mathbf{x}$ is estimated for the gold price and FTSE 100 time series. From these optimal parameter estimates we generate scenarios for both price processes assuming that the unknown noise for the next time step follows a Brownian motion. Figures 7.10 and 7.11 depict the optimal parameter estimates for the gold spot prices and for FTSE 100, respectively. Scenarios are generated starting from the last data point of the two return series. These scenarios are then used for the investment decision.

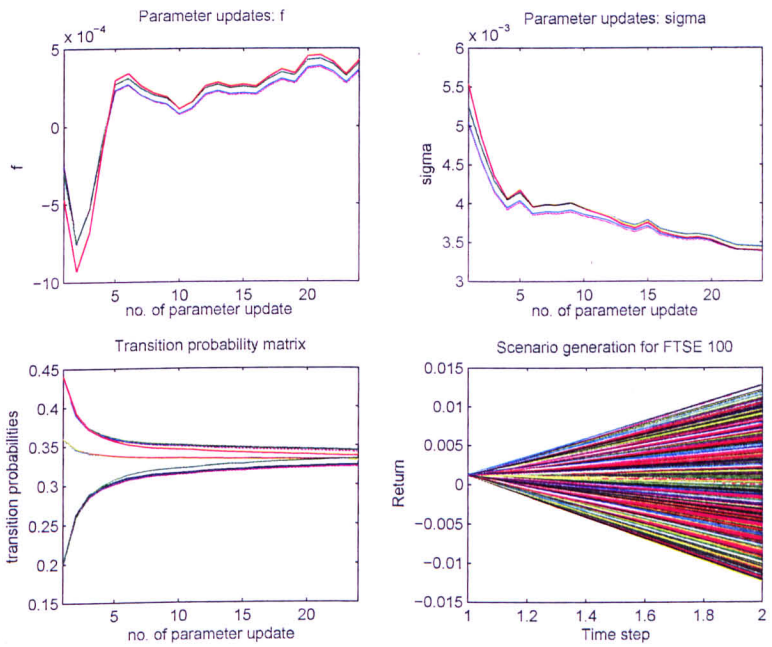


Figure 7.10: Plots of the evolution of parameter estimates and generated scenarios for FTSE 100

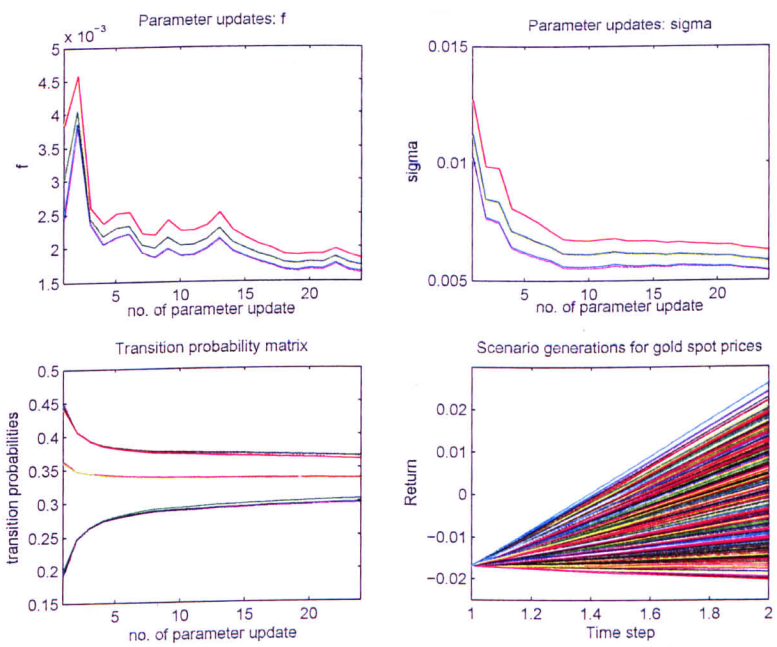


Figure 7.11: Plots of the evolution of parameter estimates and generated scenarios for gold spots

7.5.2 Scenarios in multivariate observation process setting

Scenarios are now generated under the assumption that the time series for FTSE 100 and gold spots are correlated. Due to the nature of this scenario generator, a correlation factor between the two observation processes is not included but we assume that both time series are governed by the same hidden Markov chain. Under this framework the optimal parameters of both observation processes are estimated simultaneously, the processes are assumed to be vector observations. The optimal estimate of the underlying Markov chain is filtered out from both processes. All parameters depend on the same Markov chain representing economic states which occurred over the considered time interval. When the parameters are derived for both time series, we perform the same scenario generation as discussed above. The one-step ahead scenarios are generated with the derived parameter values. Uncertainty is added by standard normal IID random variables, which are assumed to be independent for the two processes. Figures 7.12 and 7.13 below show the parameter estimation of the observation processes in this vector observation case. Possible scenarios for both time series are generated.

7.6 Portfolio optimisation with generated scenarios

This section shows the results for the portfolio optimisation based on the independent as well as the correlated scenario generations. First the solutions in the independent case for different values of a given required minimal portfolio return as well as for ten different generated scenario sets are examined. Then, this is done in the correlated case, the resulting solutions are compared with respect to the stability of the generated scenarios. The optimisation problem is solved in AMPL with the FortMP solver for a one-step time setting.

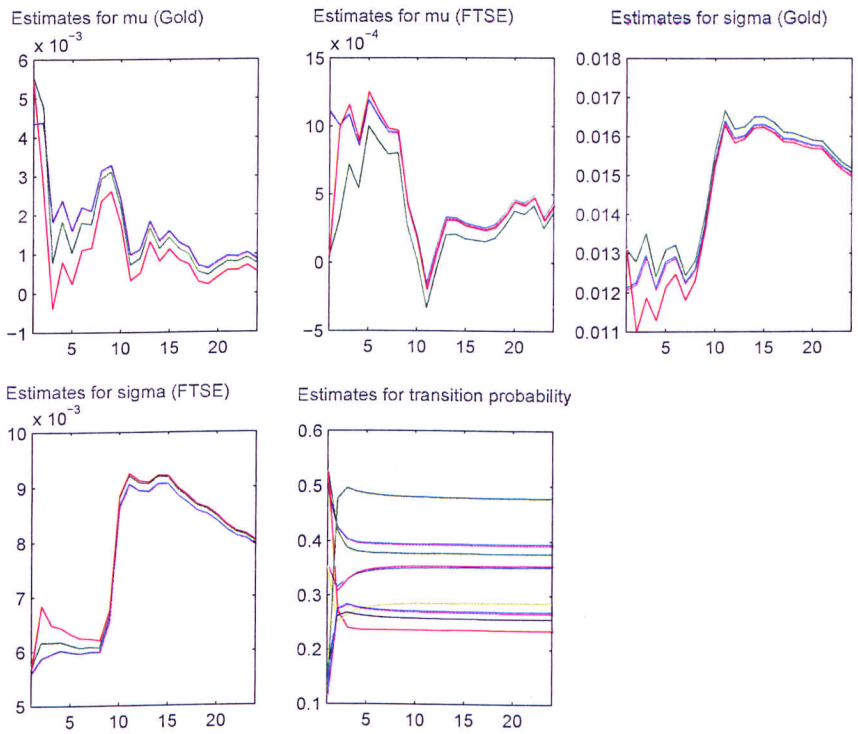


Figure 7.12: Simultaneous parameter estimation for the gold spots and FTSE 100 returns

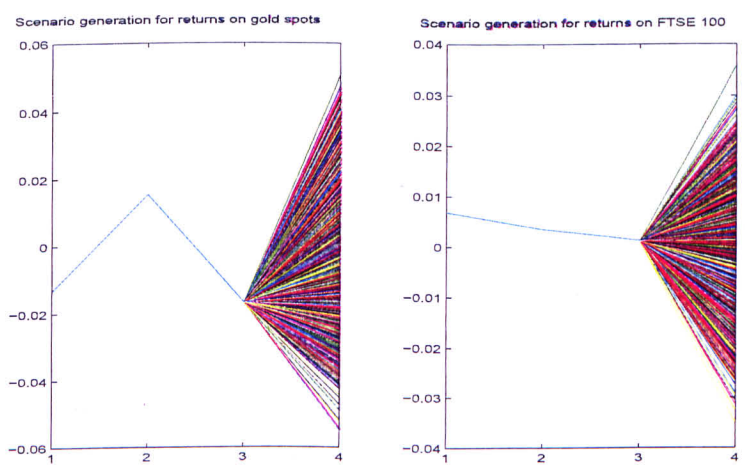


Figure 7.13: Scenarios generated for returns on gold spots and FTSE 100

7.6.1 Scenario generation for independent observation processes

Solutions for the investment problem are first obtained based on independent generated scenarios. The minimal portfolio return $MinRet$ is constant, the daily return on the bank account is assumed to be 0.0001. The confidence level of CVaR, β_{CV} is set to 1%. For both FTSE 100 and gold spot prices, the generated scenarios occur with equal probability.

The first comparison gives us the objective function and optimal portfolio weights for two different minimal portfolio returns for 2470 generated scenarios. For the first optimisation $MinRet$ is set to 0.0004. The following solution for the objective function and the percentages to be invested in gold, FTSE 100 stocks and the bank account is obtained:

Objective function:	CVaR:	0.00220126665512158
Decision variables:	BuyFTSE:	0.0710667
	BuySpots:	0.155712
	BuyBond:	0.773221

In this case, the optimal solution is to invest the largest amount of money (77.3%) in the bank account, followed by 15.6% in gold spots and 7.1% in FTSE 100 stocks. If the minimum required portfolio return is increased to $MinRet = 0.0005$, the optimal solution gives us a higher objective function. The objective and percentages invested in the different possibilities are as follows

Objective function:	CVaR:	0.0029683555401621084
Decision variables:	BuyFTSE:	0.0947556
	BuySpots:	0.207617
	BuyBond:	0.697628

The higher percentages for investing in stocks and gold spots is due to the fact that a higher minimal return is needed and therefore more risk is taken. The lower CVaR value for the lower minimal return $d = 0.0004$ reflects the lower risk of investing in a bank account.

An analysis of the scenario generation is performed on different generated scenario sets. We generate 10 sets containing 5000 scenarios each and calculate the optimal solution for these sets. The minimal required level of daily portfolio return is set to $MinRet = 0.0005$. The resulting values for the objective function as well as the the optimal percentages invested in FTSE 100 stocks, gold or bonds can be seen in Table 7.1 below. All values of the objective function lie between 0.002844 and 0.003325 with $mean = 0.0031$ and $var = 2.7070e - 008$.

Scenario set		1	2	3	4	5	6	7	8	9	10
Objective function		0.003325	0.00321	0.002955	0.003199	0.003222	0.003035	0.003004	0.002844	0.002927	0.002898
Percentages invested in:	FTSE	9.77	17.64	9.14	16.88	15.27	12.03	15.84	10.38	13.63	8.81
	Gold	24.37	20.71	22.26	21.8	22.76	23.08	21.08	21.44	21.8	22.38
	Bond	65.96	61.65	68.6	61.32	61.97	64.89	63.08	68.18	64.57	68.81

Table 7.1: Results for 10 scenario sets with independent observations

7.6.2 Scenario generation for vector observation processes

The scenario generation is performed under the assumption, that both FTSE 100 index and gold spots follow the same Markov chain. Ten sets of 10,000 scenarios are generated and the optimal solution is calculated for each of these scenario sets. The minimal required portfolio return is set to $MinRet = 0.0005$ and the CVaR confidence level is $\beta_{CV} = 0.01$. Table 7.2 shows the resulting values for the objective function as well as for the corresponding optimal portfolio weights. The value of the objective function lies between 0.015058 and 0.027259 with a mean of 0.0193 and variance of $1.4093e - 005$.

The solutions obtained with these generated scenarios are less stable than those obtained with the independent scenario generator. Their variance is a lot higher, the highest value 0.027259 is nearly twice as high as the lowest objective function value 0.015058. The investment percentages for the three different categories vary

Scenario set		1	2	3	4	5	6	7	8	9	10
Objective function		0.018185	0.015058	0.015768	0.021392	0.027259	0.016276	0.023035	0.019834	0.017423	0.0185903
Percentages invested in:	FTSE	42.47	46.13	40.58	99.47	30.78	58.77	43.06	63.75	47.06	63.1
	Gold	39.15	28.83	35.16	0.53	69.22	27.22	56.94	36.25	37.52	36.9
	Bond	18.38	25.04	24.26	0	0	14.01	0	0	15.42	0

Table 7.2: Optimisation results for 10 scenario sets with vector observations

largely between different scenario sets. In set 2, 25.04% shall be invested into the bank account compared to 0% in scenarios 4, 5, 7, 8 and 10. Compared to the objective function of the previous scenario generations, the values of the objective function are more than six times higher in the correlated setting, the variance is still low, but also significantly higher than the variance obtained by the other scenarios generated. These deviations of the objective function over different scenario sets leads to the conclusion, that this portfolio optimisation model requires independent generated scenarios. From an economic point of view, this can be explained by the largely uncorrelated price of gold to any other stock. Due to its characteristics as a stable long-term investment and a global currency, gold prices are highly likely to be independent of stock prices or indices. This is supported by stable optimal solutions achieved with independent generated scenarios.

A further analysis of generated scenario sets under different problem settings is described in the subsections 7.6.3 and 7.6.4, where one optimisation problem with maximal weight constraints is examined as well as a problem formulation for maximising the portfolio return for a given CVaR value.

7.6.3 Investment problem with weight constraints

Here, constraints on the weights invested in the different asset classes are introduced. All weights have now an upper bound of 50%, no more than half of the budget is allowed to be invested in either gold spots, FTSE 100 stocks or bonds. Taking into account the results from the previous section, this investment problem is solved with scenario sets generated by the independent scenario generator. We solve the problem with five sets containing 5,000 and five sets containing 20,000 scenarios. The objective functions obtained are again stable over the ten scenario

sets with different number of scenarios. As expected from the previous results, a

Scenario set		5,000 generated scenarios					20,000 generated scenarios				
		1	2	3	4	5	1	2	3	4	5
Objective function		0.003676	0.003541	0.003575	0.003459	0.00354	0.003522	0.003569	0.003517	0.003465	0.00352
Percentages invested in:	FTSE	29.38	32.67	32.6	31.57	30	31.84	30.55	31.87	32.04	30.05
	Gold	20.62	17.33	17.4	18.43	20	18.16	19.45	18.13	17.96	19.95
	Bond	50	50	50	50	50	50	50	50	50	50

Table 7.3: Results for the investment problem with weight constraints for 10 scenario sets

weight constraint of 50% cuts down the budget part invested in bonds from around 65% to exactly 50%. The difference is mostly invested into FTSE 100 stocks, resulting in a mean solution over ten scenarios of roughly 31% invested in FTSE 100 stocks, 19% invested in gold spots and 50% invested in the bank account. The average objective function over all generated scenario sets is 0.003538409 with a very low variance of $3.77072e-09$. The scenario generator yields sets which lead to stable solution of the investment problem with weight constraints.

7.6.4 Reformulated optimisation problem

After minimising the CVaR of the portfolio under a given minimal required level of portfolio return, the optimisation problem is now reformulated to be able to optimise the return for a given level of risk. The objective function is now to minimise the negative expected return over all scenarios, an upper bound is set for CVaR. We adopt the framework proposed in Krokhmal et.al. [96]. The percentage of the portfolio, which is allowed for risk exposure is denoted by pr and the optimisation problem is given by the objective function

$$\begin{aligned} \text{minimise} \quad & negativeReturn: \\ & -(BuyFTSE * MReturnFTSE + BuySpots * MReturnSpot \\ & \quad + BuyBond * ReturnBond) \end{aligned}$$

$$\text{subject to} \quad \text{CVaR:} \quad \alpha_V + \frac{1}{no\_scenarios * (\beta_{CV})} * \sum_{scen\ in\ SCEN} (negdev[scen]) \leq pr\ .$$

First, the optimal solution is obtained on one scenario set with 5000 scenarios. An efficient frontier can be calculated by setting varying values of pr. The efficient

frontier is plotted in Figure 7.14 for three β_{CV} values: $\beta_{CV} = 0.1, \beta_{CV} = 0.05$ and $\beta_{CV} = 0.01$.

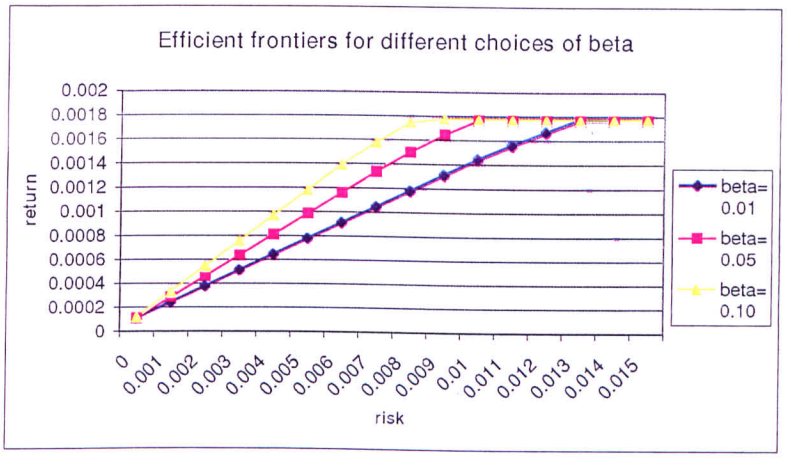


Figure 7.14: Efficient frontier for the portfolio selection problem

Again, scenario sets are generated with the generation method for independent observations. We generate five sets with 5,000 scenarios and five with 10,000 scenarios to test the in-sample stability of the scenario generator. Table 7.4 shows the results for maximising the portfolio return for a given level of $\beta_{CV} = 0.01$. and $pr = 0.005$. A weight constraint $\leq 50\%$ is also included in this optimisation problem.

Scenario set		5000 generated scenarios					10000 generated scenarios				
		1	2	3	4	5	1	2	3	4	5
Objective function		-0.000732	-0.000802	-0.000771	-0.00076	-0.000672	-0.000715	-0.000709	-0.00076	-0.000721	-0.000741
Percentages invested in:	FTSE	19.51	27.38	22.81	19.49	20.65	15.86	16.79	17.21	24.84	21.32
	Gold	35.03	34.34	34.81	35.57	34.01	35.72	34.8	36.47	33.16	34.18
	Bond	45.46	38.28	42.38	44.94	45.34	48.42	48.41	46.32	42	44.51

Table 7.4: Results for the investment problem with minimised negative return

The solutions for this problem are quite stable throughout the scenario sets. The mean value of the objective function is -0.000738382 with corresponding mean percentages of 20.59% to invest in FTSE 100 stocks, 34.81% to invest in gold and 44.60% to put into a bank account. The variance of the objective function, which is $1.35182e-09$, is very low.

7.7 Some concluding remarks

An asset allocation problem can be examined through a variety of techniques. In this chapter the HMM framework is utilised for asset allocation problem in two different settings. In the first part of this chapter a filtering method for a multivariate observation process was considered. Using the forecasted returns as a signal for an investment decision, an investment strategy that allows switching between investments in growth or value stocks according to the risk-adjusted signals received through the HMM filters is analysed. As expected, this switching strategy does not always lead to a higher investment return than the pure strategies. However, the HMM signals offer a good guidance for switching times, when a decision has to be made on when to invest in growth or value assets. Furthermore optimal expected weights are calculated through a mean-variance utility function. These optimal weights depend on the parameter forecast for both processes in the vector observation. A mixed investment strategy, where the optimal asset allocation depends on the optimal calculated weights is able to reduce the risk compared to simply following a pure NASDAQ or Dow Jones investment strategy.

In the second part of this chapter a scenario generation technique within the HMM framework was developed. Here, parameter estimates obtained through the filters for processes of the Markov chain were employed to generate scenarios of asset price processes. With these scenario sets various investment problems under a CVaR framework were solved. Scenarios generated in a one-dimensional observation process setting led to stable solutions of the optimisation problem. Scenarios generated on the basis of vector observations assume that both time series follow one underlying Markov chain and are therefore dependent on the same economic triggers. These scenario sets lead to less stable results, which supports the fact that the evolution of gold spots is widely independent of the FTSE 100. Generated scenarios were also tested on the investment problem including weight constraints on the percentages invested in different assets. Furthermore the CVaR optimisation problem was reformulated so that the daily return is maximised subject to a given level of risk. This problem was solved with ten different scenario sets having 5,000 or 10,000 scenarios each. Again, the generated scenarios lead to stable solutions.

The HMM scenario generator is therefore able to provide scenario sets for a variety of investment problems. The focus of future research will be to extend and test the scenario generator for multiple time step optimisation problems.

Chapter 8

Pricing of credit default swaps in an HMM setting

The recent turbulence on financial markets around the world sharply increased the concerns of investors about credit quality and valuation of structured credit products. Although the origin of this credit crisis in more recent times can be attributed in the most part to the subprime mortgage market, other financial markets were highly affected and questions are raised about the accuracy of evaluating structured products and the risks associated with them. The trading activities in the credit markets nowadays are comparable to those in the traditional markets of bond, equity and foreign exchange. The large impacts on the financial industry caused by the turmoil on the subprime mortgage market raised the need for reliable and realistic valuations of credit risk. Credit default swaps (CDS) are amongst the most popular types of credit derivative contracts being traded actively in the credit markets around the world. The valuation of credit default swaps is performed here under an extended version of Merton's structural model for firm's corporate liabilities. In this modelling framework we assume that the Markov chain is observable. The interest rate process of a money market account, the appreciation rate and the volatility of the firm's value are governed by this Markov chain in continuous time with finite number of states, where the states of the Markov chain represent states of the economy. Certain factors which affect the value of the firm, e.g. changes in business conditions, company operations, management decisions or macroeconomic conditions lead to regime-switching of the parameters. A description of credit de-

fault swaps and an overview of the framework is given in section 8.1. In section 8.2 the Esscher transform is employed to determine an equivalent martingale measure for the valuation problem in incomplete markets. Section 8.3 discusses the derivation of systems of coupled partial differential equations (PDEs) satisfied by the real-world and risk-neutral default probabilities for valuating the swap rate of a CDS. In section 8.4 the regime-switching effect of the firm's values are investigated via a numerical approximation for the case of a 2-state Markov chain. A sensitivity analysis for different model parameters is performed for the real-world default probability and the swap rate in section 8.5. The numerical results of the PDE approach are compared to those from a Monte Carlo simulation in section 8.6.

8.1 Credit default swaps

A CDS is a financial contract involving two parties, namely, the protection buyer and the protection seller. The protection buyer pays the protection seller periodic premiums until the maturity of the contract or occurrence of a credit event, such as the default of a reference asset or entity. The protection seller pays the loss incurred by the protection buyer due to the credit event. In other words, the protection buyer secures some form of protection (similar to an insurance contract) against the impact of a credit event involving the underlying reference asset or entity, such as a corporate bond or a loan, from the protection seller. When a credit event occurs, the protection buyer also needs to pay a final accrual fee to the protection seller. In return to all the paid premiums and accrual fee, the protection seller pays the protection buyer an amount equal to the CDS notional amount face value minus the recovery value at the time when the reference entity defaults.

In the analysis of a CDS contract, it is necessary to determine the swap rate. The swap rate of a CDS is calculated by equating the sum of the expected present values of the fee leg, which consists of the periodic premium payments to the sum of the expected present values of the contingent payment leg which consists of the payments in case of a credit event. The default probabilities computed under a risk-neutral measure are the key elements in the CDS swap rate computation.

There are two major strands of literature concerning the valuation of credit products. The first one is based on structural credit risk models, which are also known as the Merton structural firm model or the Merton default model, pioneered by the seminal works of Black and Scholes [22] and Merton [115], which are then extended by Black and Cox [20], Longstaff and Schwartz [104], Leland [100] and Zhou [150], amongst others. In these models, the market value of a firm is assumed to be governed by a diffusion process and the occurrence of a default is triggered by the event that the market value of the firm drops below a certain threshold level, called the default barrier. In structural models, the relationship between the capital structure of a firm and its default is modelled explicitly and endogenously. Another strand of literature is based on the reduced-form credit risk models developed in Artzner and Delbaen [5], [6], [7] Jarrow and Turnbull [92], Lando [98], Madan and Unal [108] and Duffie and Singleton [47], amongst others. In the reduced-form credit risk models, the occurrences of defaults are modelled exogenously. In particular, the occurrence of a default is modelled by a random point process and the time of default is described by a stopping time, which is unpredictable with respect to the information generated by the market value of a firm. For a detailed discussion of the Merton default model and the reduced-form credit risk models see Lando [99]. The approach in this chapter is an extension of the Merton structural firm value model based on the Merton structural model. Specifically, we develop a Markov-modulated version of the Merton structural firm model for the valuation of CDS. In the standard approach of the Merton structural firm model, default occurs when the firm's value drops below the face value of the debt at the maturity of the debt. In the first-passage-time approach, default occurs at any time during the lifetime of the debt once the firm's value drops below a default barrier level. The first-passage-time approach, which is employed here offers a more flexible way than the standard approach in describing the occurrence of default. The interest rate at which a money market account grows, the appreciation rate and volatility of the firm's values have switching dynamics; they all follow the evolution of a finite state continuous-time Markov chain. In this model the Markov chain is assumed to be observable, the states could represent the states of an economy. The economic intuition behind the Markov-modulated Merton structural firm model is to incorporate the impact of the transitions of macroeconomic conditions and dif-

ferent states of business cycles on the market value of a firm via the introduction of the Markovian regime-switching principle. In a recent paper by Hackbarth et al. [77] the impact of macroeconomic conditions on credit risk and dynamic capital structure is studied. They find that incorporating the impact of macroeconomic conditions and business cycles via the introduction of Markovian regime-switching effect has important empirical implications for corporations and explains some important empirical characteristics of observed credit spreads. During the periods of economic expansion, internal consumptions and demands of goods or services remain strong. Hence, firms can generate more business and earn higher profits. This makes the expected rates of growth of the firms' values or the appreciation rates higher. During the periods of economic recession, internal consumptions are weak. This worsens the earnings of the firms and makes the profits more uncertain. In this case, it is expected that the appreciation rates of the firms' values will be lower and the volatilities of the firms' values will be higher. The Markov-modulated version of the Merton structural model considered here can describe these realistic situations.

The use of a regime-switching model intends to improve the Merton structural model by including jumps or discontinuities in the market value of the firm through the possibility of regime shifts. There is strong empirical evidence that Merton-based structural models underestimate the actual probability of default. Tarashev [142] found that the actual probability of default obtained by the Merton structural model is significantly less than the empirical default rate. Leland [101] mentioned that actual short-term default probabilities tend to be underestimated by structural models. Since the firm's value is described by a diffusion process in the Merton structural model, it does not allow sudden jumps in the firm's value, and hence defaults can never happen by surprise. This is one possible reason why the actual short-term probabilities of defaults are underestimated by the Merton structural model. Leland [101] suggested that the underestimation of the actual default probabilities can be improved by including a jump component in the dynamics of the firm's value. Zhou [151] incorporated jump risk into the default process in modelling the term structure of credit spreads and describes the firm's value as a jump-diffusion model. This regime-switching model discussed here is another

possible way to include discontinuities in the market value of the firm. Transitions of macroeconomic conditions or different stages of business cycles are mirrored in switching parameters. The market value of the firm is modelled here through a two-state Markovian regime-switching Merton structural model. As in the Merton structural model, we consider a frictionless and competitive continuous-time market in which the following assumptions hold: (i) There are no transaction costs. (ii) Agents are price-takers. (iii) Assets are divisible; that is, assets can be traded in any fractional units. (iv) Short selling is allowed. (v) The firm pays no dividends.

Let $\mathbf{x} := \{\mathbf{x}_t, t \geq 0\}$ denote a finite-state, continuous-time Markov chain. We have the following semimartingale form for the dynamics of the continuous-time Markov chain $\mathbf{x}$ as in Elliott et al. [49]

$$\mathbf{x}_t = \mathbf{x}_0 + \int_0^t \mathbf{Q}^\top \mathbf{x}_u du + \mathbf{v}_t \ , \tag{8.1}$$

Here, $\mathbf{v}_t$ is a martingale with respect to the sigma field generated by the process $\mathbf{x}$ and $\mathbf{Q} := [q_{ij}]_{i,j=1,2,\ldots,N}$ denotes the rate matrix or generator of the chain $\mathbf{x}$, where q_{ij} is the rate at which $\mathbf{x}$ makes a transition into state j when it is in state i. q_{ij}, $i, j = 1, 2, \ldots, N$, are called the instantaneous transition rates.
Without loss of generality, we assume that the Markov chain $\mathbf{x}$ takes values from the set of unit vectors in $\mathbb{R}^N$, "$\{\mathbf{x}_t = \mathbf{e}_i\}$" represents the event that the economy is in state i at time t. The Markov chain is observable and serves as a proxy for some exogenous economic factors, such as Gross Domestic Product (GDP) and price indices. For a two-state Markovian regime-switching Merton default model, a continuous-time Markov chain $\mathbf{x}$ with two states (i.e. , $N = 2$) is considered, where state "1" (state "2") represents a "Good" ("Bad") economic state.

Let $\breve{r}_t$ denote the instantaneous interest rate of a money market account at time $t \geq 0$. Then, the instantaneous interest rate of the money market account is given by:

$$\breve{r}_t = \langle \breve{\mathbf{r}}, \mathbf{x}_t \rangle \ , \tag{8.2}$$

where $\breve{\mathbf{r}} := (\breve{r}_1, \breve{r}_2, \ldots, \breve{r}_N)$ with $\breve{r}_i > 0$ for each $i = 1, 2, \ldots, N$. So, if the economy is in state i at time t (i.e. $\mathbf{x}_t = \mathbf{e}_i$), $\breve{r}(t) = \langle \breve{\mathbf{r}}, \mathbf{x}_t \rangle = \breve{r}_i$. Let μ_t and ς_t denote

the expected growth rate and the volatility rate of the firm's value, respectively, at time $t \geq 0$. Similarly, the expected rate of growth and volatility rate of the firm's value are:

$$\mu_t = \langle \boldsymbol{\mu}, \mathbf{x}_t \rangle \ , \tag{8.3}$$

$$\varsigma_t = \langle \boldsymbol{\varsigma}, \mathbf{x}_t \rangle \ , \tag{8.4}$$

where $\boldsymbol{\mu} := (\mu_1, \mu_2, \ldots, \mu_N)$ and $\boldsymbol{\varsigma} := (\varsigma_1, \varsigma_2 \ldots, \varsigma_N)$ with $\mu_i > \check{r}_i$ and $\varsigma_i > 0$ for each $i = 1, 2, \ldots, N$.

Let $W := \{W_t, t \geq 0\}$ denote a standard Brownian motion under a real-world probability $\mathcal{P}$. We assume that W and $\mathbf{x}$ are stochastically independent under $\mathcal{P}$. Let $\{V_t, t \geq 0\}$ denote the firm's value process. Then, the two-state Markovian regime-switching Merton default model for the firm's value process is given by:

$$dV_t = \mu_t V_t dt + \varsigma_t V_t dW_t \tag{8.5}$$

whilst keeping in mind the Markov chain specification of the drift and volatility in (8.3) and (8.4). Here, following some literature on the Merton default model (see, for example, Merton [115], Black and Cox [20] and Longstaff and Schwartz [104] we assume that the value of the firm can be directly observable. By Itô's lemma,

$$V_t = V_0 \exp\left[\int_0^t \left(\mu_u - \frac{1}{2}\varsigma_u^2\right) du + \int_0^t \varsigma_u dW_u\right] . \tag{8.6}$$

8.2 The Esscher transform and swap rates

Due to the additional uncertainty brought about by the switching of regimes, the market is no longer complete. Hence, there are infinitely many equivalent martingale measures for valuation. Guo [76] introduced Arrow-Debreu securities related to the cost of switching to complete the market, which could hedge away the additional uncertainty induced by switching regimes. This approach here will make use of the regime-switching Esscher transform to find an equivalent martingale measure. The regime-switching Esscher transform was adopted in Elliott et al. [50] for option valuation in the framework of a Markov-modulated Black-Scholes-Merton

market. Let $\mathcal{T}$ denote the time index of the model. Without any dividend payments, suppose that $U_t := \ln(V_t/V_0)$ is the continuously compounded return of the firm's value V over the interval $[0,t]$, $t \in \mathcal{T}$. We shall define the regime-switching Esscher transform on the basis of the variable U_t. For each $t \in \mathcal{T}$, let $\mathcal{F}_t^U$ denote the information set generated by the value of the process $\mathbf{U} := \{U_t\}_{t\in\mathcal{T}}$ up to and including time t. For each $t \in \mathcal{T}$, let $\mathcal{G}_t$ denote the information set generated by the values of the processes $\mathbf{x}$ and $\mathbf{U}$ up to and including time t. Moreover, define a real-valued stochastic process $\{\vartheta_t\}_{t\in\mathcal{T}}$ as

$$\vartheta_t := \vartheta(t, \mathbf{x}_t) = \langle \boldsymbol{\vartheta}, \mathbf{x}_t \rangle \ , \tag{8.7}$$

where $\boldsymbol{\vartheta} := (\vartheta_1, \vartheta_2, \ldots, \vartheta_N)^\top \in \mathbb{R}^N$. As in Elliott et al. [50], the regime-switching Esscher transform $\mathcal{P}^\vartheta$ equivalent to $\mathcal{P}$ on $\mathcal{G}_t$ associated with the process $\{\vartheta(t)\}_{t\in\mathcal{T}}$ in (8.7) is given by

$$\mathrm{E}\left[\left.\frac{d\mathcal{P}^\vartheta}{d\mathcal{P}}\right| \mathcal{G}_t\right] = \frac{\exp\left(\int_0^t \vartheta_s dU_s\right)}{\mathrm{E}\left[\left.\exp\left(\int_0^t \vartheta_s dU_s\right)\right| \mathcal{F}_t^{\mathbf{x}}\right]} \ , \quad t \in \mathcal{T} \ , \tag{8.8}$$

where $\mathrm{E}[\cdot]$ denotes expectation taken with respect to $\mathcal{P}$. Note further that the Radon-Nikodým derivative defined in (8.8) of the regime-switching Esscher transform can be expressed as

$$\mathrm{E}\left[\left.\frac{d\mathcal{P}^\vartheta}{d\mathcal{P}}\right| \mathcal{G}_t\right] = \exp\left(\int_0^t \vartheta_s \varsigma_s dW_s - \frac{1}{2}\int_0^t \vartheta_s^2 \varsigma_s^2 ds\right) . \tag{8.9}$$

The fundamental theorem of asset pricing (see Harrison and Kreps [86], Harrison and Pliska [87], [88] and Delbaen and Schachermayer [41]) states that the absence of arbitrage opportunities is "essentially" equivalent to the existence of an equivalent martingale measure under which the discounted price process of a non-dividend-paying stock is a martingale. Let $\tilde{\mathcal{G}}_t := \mathcal{F}_T^{\mathbf{x}} \vee \mathcal{F}_t^U$, for each $t \in \mathcal{T}$. Write $\tilde{V}_t := \exp(-\int_0^t \check{r}_t du) V_t$, which denotes the discounted firm's value at time $t \in \mathcal{T}$. Then, the martingale condition in this case is defined with respect to an enlarged information structure $\tilde{\mathcal{G}}$ and gives

$$\tilde{V}_u = \mathrm{E}^\vartheta[\tilde{V}_t | \tilde{\mathcal{G}}_u] \ , \quad \mathcal{P}\text{-almost surely, for any } t, u \in \mathcal{T} \text{ with } t \geq u \ ,$$

where $\mathrm{E}^{\vartheta}[\cdot]$ denotes an expectation with respect to $\mathcal{P}^{\vartheta}$. In particular, by setting $u = 0$, we require that for each $t \in \mathcal{T}$, the random variable

$$\mathrm{E}^{\vartheta}\left[\exp\left(-\int_0^t \breve{r}_s ds + U_t\right)\middle|\mathcal{F}_t^{\mathbf{x}}\right] = 1, \quad \mathcal{P}\text{-almost surely}\ .$$

This condition is given in Elliott et al. [50]. In the case when there is no regime-switching, the condition becomes the one described in Gerber and Shiu [71].

Elliott et al. [50] demonstrate that the martingale condition implies that, for each $t \in \mathcal{T}$,

$$\vartheta_t = \frac{\breve{r}_t - \mu_t}{\varsigma_t^2}\ . \tag{8.10}$$

From equation (8.10), the Radon-Nikodým derivative in (8.9) becomes

$$\mathrm{E}\left[\frac{d\mathcal{P}^{\vartheta}}{d\mathcal{P}}|\mathcal{G}_t\right] = \exp\left(\int_0^t \left(\frac{\breve{r}_s - \mu_s}{\varsigma_s}\right) dW_s - \frac{1}{2}\int_0^t \left(\frac{\breve{r}_s - \mu_s}{\varsigma_s}\right)^2 ds\right). \tag{8.11}$$

Invoking Girsanov's theorem with the aid of (8.11), $W_t^{\vartheta} = W_t - \int_0^t \left(\frac{\breve{r}_s - \mu_s}{\varsigma_s}\right) ds$ is a standard Brownian motion with respect to $\{\mathcal{G}_t\}_{t\in\mathcal{T}}$ under $\mathcal{P}^{\vartheta}$. Here, we also assume that W and $\mathbf{x}$ are stochastically independent under $\mathcal{P}^{\vartheta}$. Under this assumption, the probability law of the Markov chain $\mathbf{x}$ remains unchanged when changing the probability measures from $\mathcal{P}$ to $\mathcal{P}^{\vartheta}$. Consequently, from (8.5) we can write the market value of the firm conditional on $\mathcal{G}_t$ under $\mathcal{P}^{\vartheta}$ in terms of W_t^{ϑ} as

$$dV_t = \breve{r}_t V_t dt + \varsigma_t V_t dW_t^{\vartheta}\ . \tag{8.12}$$

Here, as in the Longstaff and Schwartz [104] model, an "exogenous" default boundary is considered which is assumed to be a given constant.

Let ϱ be the first passage time of default. That is,

$$\varrho := \inf\{t \in [0,T] | V_t \leq L\}\ . \tag{8.13}$$

Incorporating (8.13), this means that the risk-neutral default probability under $\mathcal{P}^{\vartheta}$ is given by

$$\mathcal{P}^{\vartheta}(\varrho \leq T) = \mathcal{P}^{\vartheta}\left(L \geq \min_{t\in[0,T]} V_t\right)\ . \tag{8.14}$$

For the determination of the swap rate (SR) of a CDS, we equate the sum of the expected present values of the fee leg and the expected present values of the contingent payment leg linked to a credit event. Suppose that $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$ and that SR is paid by the protection buyer to the protection seller at dates $t_1 < t_2 < \cdots < t_n$ if the default has not yet happened. Let $D(t_{k-1}, t_k)$ denote the interval between payment dates. For example, when there are semi-annual payments, $D(t_{k-1}, t_k) = 0.5$, for each $k = 1, 2, \ldots, n$. Suppose $\{\mathcal{P}^{\vartheta}(t_k) | k = 1, 2, \ldots, n\}$ constitutes the term structure of risk-neutral default probabilities over different payment points. Then, the sum of the expected present values of the fee leg, denoted by FL, is

$$\begin{aligned} FL &= SR \sum_{k=1}^{n} D(t_{k-1}, t_k)(1 - \mathcal{P}^{\vartheta}(t_k)) \exp\left(- \int_0^{t_k} \langle \breve{\mathbf{r}}, \mathbf{x}_u \rangle \, du \right) \\ &:= SR \times KL \,, \end{aligned} \tag{8.15}$$

where $\mathcal{P}^{\vartheta}(t_k) := \mathcal{P}^{\vartheta}(\varrho \leq t_k)$, for each $k = 1, 2, \ldots, n$, is explicitly given by (8.14) and KL is defined as

$$KL := \sum_{k=1}^{n} D(t_{k-1}, t_k)(1 - \mathcal{P}^{\vartheta}(t_k)) \exp\left(- \int_0^{t_k} \langle \breve{\mathbf{r}}, \mathbf{x}_u \rangle \, du \right) .$$

In practice, it is assumed that default between regular fee payments usually occurs halfway over the period. With this assumption, the sum of the expected present values of fee accruals, denoted by AL, is given by

$$\begin{aligned} AL &= SR \sum_{k=1}^{n} \frac{D(t_{k-1}, t_k)}{2} (\mathcal{P}^{\vartheta}(t_k) - \mathcal{P}^{\vartheta}(t_{k-1})) \exp\left(- \int_0^{t_k} \langle \breve{\mathbf{r}}, \mathbf{x}_u \rangle \, du \right) \\ &:= SR \times ML \,, \end{aligned} \tag{8.16}$$

where

$$ML := \sum_{k=1}^{n} \frac{D(t_{k-1}, t_k)}{2} (\mathcal{P}^{\vartheta}(t_k) - \mathcal{P}^{\vartheta}(t_{k-1})) \exp\left(- \int_0^{t_k} \langle \breve{\mathbf{r}}, \mathbf{x}_u \rangle \, du \right) .$$

Of course, the accuracy of the approximation is improved if the time steps between regular payments are made smaller.

Let RR denote the fixed constant rebate of the underlying reference asset or entity. Then, the expected payoff coming from the contingent payment leg is given by

$$PL = (1 - RR) \sum_{k=1}^{n} (\mathcal{P}^{\vartheta}(t_k) - \mathcal{P}^{\vartheta}(t_{k-1})) \exp\left(- \int_0^{t_k} \langle \breve{\mathbf{r}}, \mathbf{x}_u \rangle \, du \right) . \tag{8.17}$$

The results from equations (8.15), (8.16) and (8.17), combined altogether, tell us that

$$PL = FL + AL = SR(KL + ML) \ . \tag{8.18}$$

Hence, from (8.18) it is easily seen that the swap rate SR can be expressed as

$$SR = \frac{PL}{KL + ML} \ . \tag{8.19}$$

For the valuation of credit derivatives, risk-neutral default probabilities have to be calculated, which reflect the current market expectations on future default probabilities. The real-world default probability reflects the "true" or actual likelihood of defaults. In the situation when there is no regime switching, exact formulae for default probabilities can be obtained by using formula D.2 in Gerber and Shiu [72] (see also lemma 7.8.5 of Elliott and Kopp [54] and Corollary B.3.1 of Musiela and Rutkowski [120]). In this case, the risk-neutral default probability is given by

$$\mathcal{P}^{\vartheta}(t_k|V) = 1 - \Phi(d_1) + \left(\frac{L}{V}\right)^{\frac{2\breve{r}}{\varsigma^2} - 1} \Phi(d_2) \ , \tag{8.20}$$

where

$$d_1 = \frac{\ln(V/L) + (\breve{r} - \frac{1}{2}\varsigma^2)t_k}{\varsigma\sqrt{t_k}} \ ,$$

$$d_2 = \frac{\ln(L/V) + (\breve{r} - \frac{1}{2}\varsigma^2)t_k}{\varsigma\sqrt{t_k}} \ ,$$

and $\Phi(x)$ represents the cumulative distribution function of a standard normal random variable.

The only difference between the real-world default probabilities and the risk-neutral

default probabilities is that μ_i's $(i = 1, 2, \ldots, N)$ are used in the real-world default probabilities while $\breve{r}_i$'s $(i = 1, 2, \ldots, N)$ are involved in the risk-neutral default probabilities. So, the real-world default probability is given by:

$$\mathcal{P}^{\vartheta}(t_k|V) = 1 - \Phi(d_1) + \left(\frac{L}{V}\right)^{\frac{2\mu}{\varsigma^2}-1} \Phi(d_2) \,, \tag{8.21}$$

where

$$\begin{aligned} d_1 &= \frac{\ln(V/L) + (\mu - \frac{1}{2}\varsigma^2)t_k}{\varsigma\sqrt{t_k}} \,, \\ d_2 &= \frac{\ln(L/V) + (\mu - \frac{1}{2}\varsigma^2)t_k}{\varsigma\sqrt{t_k}} \,. \end{aligned}$$

A system of coupled PDEs for the default probabilities in the non-regime-switching as well as the regime-switching case is derived in section 8.3. The PDEs are numerically approximated to solve the default probabilities and swap rates. These numerical results are furthermore compared to default probabilities obtained through the Monte-Carlo simulation.

8.3 System of coupled PDEs for the default probabilities

Consider first the risk-neutral default probabilities under the risk-neutral probability measure $\mathcal{P}^{\vartheta}$. Similar arguments can be applied to derive a system of coupled PDEs satisfied by the real-world default probabilities under $\mathcal{P}$. For each $k = 1, 2, \ldots, n$, and $t \leq t_k$, define the conditional risk-neutral default probability $\mathcal{P}^{\vartheta}(\varrho \leq t_k|\mathcal{G}_t)$. Let $H_k := I_{\{L \geq \min_{t\in[0,t_k]} V_t\}}$, whose value is known if $\mathcal{G}_{t_k}$ is given. Write $\underline{m}(t) = \min_{u\in[0,t]} V_u$. Then,

$$\begin{aligned} \mathcal{P}^{\vartheta}(\varrho \leq t_k|\mathcal{G}_t) &= \mathrm{E}^{\vartheta}(H_k|\mathcal{G}_t) \\ &= \mathrm{E}^{\vartheta}(H_k|\mathbf{x}_t = x, V_t = v, \underline{m}(t) > L) \,, \end{aligned} \tag{8.22}$$

since $\mathbf{x}$ and V are Markov processes with respect to $\mathcal{G}$.

Let $\tilde{f}(t,x,v) := \mathrm{E}^{\vartheta}(H_k|\mathbf{x}_t = x, V_t = v, \underline{m}(t) > L)$ where the conditional expectation must be evaluated under $\mathcal{P}^{\vartheta}$ as specified in (8.22). Write $\tilde{f}_t^i := \tilde{f}(t,\mathbf{e}_i,v)$ for $i = 1,2,\ldots,N$ and $\tilde{\mathbf{f}} := (\tilde{f}^1,\tilde{f}^2,\ldots,\tilde{f}^N)^\top$. Then, by Itô's differentiation rule and the representation of $\mathbf{x}_t$ in Buffington and Elliott [28] or equation (3.3), in the region $\{\underline{m}(t) > L\}$,

$$\begin{aligned}
&\tilde{f}(t,x,v)\\
&= \tilde{f}(0,x_0,v_0) + \int_0^t \frac{\partial \tilde{f}}{\partial u}du + \int_0^t \frac{\partial \tilde{f}}{\partial v}dV_u + \frac{1}{2}\int_0^t \frac{\partial^2 \tilde{f}}{\partial v^2}\varsigma_u^2 V_u^2 du\\
&\quad + \int_0^t \left\langle \tilde{\mathbf{f}}, d\mathbf{x}_u \right\rangle\\
&= \tilde{f}(0,x_0,v_0) + \int_0^t \left(\frac{\partial \tilde{f}}{\partial u} + r_u V_u \frac{\partial \tilde{f}}{\partial v} + \frac{1}{2}\frac{\partial^2 \tilde{f}}{\partial v^2}\varsigma_u^2 V_u^2\right)du\\
&\quad + \int_0^t \left\langle \tilde{\mathbf{f}}, \mathbf{Q}^\top \mathbf{x}_u \right\rangle du + \int_0^t \frac{\partial \tilde{f}}{\partial v}\varsigma(u)V_u dW_u + \int_0^t \left\langle \tilde{\mathbf{f}}, d\mathbf{v}_u \right\rangle . \qquad (8.23)
\end{aligned}$$

Since $\tilde{f}(t,x,v) = \mathrm{E}^{\theta}(H_k|\mathcal{G}_t)$ is a $(\mathcal{G},\mathcal{P}^{\theta})$-martingale, all the bounded variation terms of the stochastic integral of $\tilde{f}(t,x,v)$ in (8.23) must sum to zero. So, for each $i = 1,2,\ldots,n$ and $t \le t_k$, $\tilde{f}$ satisfies the PDE in the region $\{\underline{m}(t) > L\}$

$$\frac{\partial \tilde{f}}{\partial t} + \breve{r}_t V_t \frac{\partial \tilde{f}}{\partial v} + \frac{1}{2}\frac{\partial^2 \tilde{f}}{\partial v^2}V_t^2\varsigma_t^2 + \left\langle \tilde{\mathbf{f}}, \mathbf{Q}^\top \mathbf{x}_t \right\rangle = 0 , \qquad (8.24)$$

with boundary conditions

$$\tilde{f}(t,x,L) = 1 , \quad \tilde{f}(t,x,\infty) = 0 , \qquad (8.25)$$

and terminal condition

$$\tilde{f}(t_k,x,v) = H_k = I_{\{L \ge \min_{t\in[0,t_k]} V_t\}} . \qquad (8.26)$$

Since $\mathbf{x}$ takes one of the $\mathbf{e}_i$'s for $i = 1,2,\ldots,N$ and $t \le t_k$, in the region $\{\underline{m}(t) > L\}$ equations (8.24), (8.25) and (8.26) correspond, respectively, to

$$\frac{\partial \tilde{f}^i}{\partial t} + \breve{r}^i V_t \frac{\partial \tilde{f}^i}{\partial v} + \frac{1}{2}\frac{\partial^2 \tilde{f}^i}{\partial v^2}(\varsigma^i)^2 V_t^2 + \left\langle \tilde{\mathbf{f}}, \mathbf{Q}^\top \mathbf{e}_i \right\rangle = 0 , \qquad (8.27)$$

with boundary conditions

$$\tilde{f}(t,\mathbf{e}_i,L) = 1 , \quad \tilde{f}(t,\mathbf{e}_i,\infty) = 0 , \qquad (8.28)$$

and terminal condition

$$\tilde{f}(t_k, \mathbf{e}_i, v) = H_k = I_{\{L \geq \min_{t\in[0,t_k]} V_t\}} , \tag{8.29}$$

for each $i = 1, 2, \ldots, N$.

Now, let $f(t, x, v) := \mathrm{E}(H_k|\mathcal{G}(t))$, which is the conditional real-world default probability. Write $f^i := f(t, e_i, v)$, for each $i = 1, 2, \ldots, N$ and $\mathbf{f} := (f^1, f^2, \ldots, f^N)^\top \in \mathbb{R}^N$. Then, by noticing that $f(t, x, v) = \mathrm{E}(H_k|\mathcal{G}_t)$ is a $(\mathcal{G}_t, \mathcal{P})$-martingale and applying the similar argument as above, the following system of coupled PDEs for the real-world default probabilities f^i $(i = 1, 2, \ldots, N)$ for different states can be derived in the region $\{\underline{m}(t) > L\}$ and they are presented as follows:

$$\frac{\partial f^i}{\partial t} + \mu^i V(t)\frac{\partial f^i}{\partial v} + \frac{1}{2}\frac{\partial^2 f^i}{\partial v^2}(\varsigma^i)^2 V_t^2 + \langle \mathbf{f}, \mathbf{Q}^\top \mathbf{e}_i \rangle = 0 , \tag{8.30}$$

with boundary conditions

$$f(t, \mathbf{e}_i, L) = 1 , \quad f(t, \mathbf{e}_i, \infty) = 0 , \tag{8.31}$$

and terminal condition

$$f(t_k, \mathbf{e}_i, v) = H_k = I_{\{L \geq \min_{t\in[0,t_k]} V_t\}} , \tag{8.32}$$

for each $i = 1, 2, \ldots, N$.

8.4 Numerical approximation

The numerical approximation to the solution of the system of coupled PDEs (8.30)-(8.32) for the real-world default probabilities is discussed in this section. A two-state Markov chain $\mathbf{x}$ with a rate matrix $\mathbf{Q} = [q_{ij}]_{i,j=1,2}$ is considered, where $-q_{11} = q_{12}$ and $q_{21} = -q_{22}$. Note that for the numerical approximation to risk-neutral default probabilities under the measure $\mathcal{P}^\theta$, the drift (μ^1, μ^2) in the system of coupled PDEs under the real-world measure $\mathcal{P}$ has to be changed to $(\breve{r}^1, \breve{r}^2)$, where $\breve{r}^1 < \mu^1$ and $\breve{r}^2 < \mu^2$.

Here, the real-world default probabilities (f^1, f^2) corresponding to the term t_k satisfy the following system of two coupled PDEs in the region $\{\underline{m}(t) > L\}$

$$\frac{\partial f^1}{\partial t} + \mu^1 V_t \frac{\partial f^1}{\partial v} + \frac{1}{2}\frac{\partial^2 f_1}{\partial v^2}(\varsigma^1)^2 V^2(t) - q_{11} f^1 + q_{22} f^2 = 0 \,, \tag{8.33}$$

$$\frac{\partial f^2}{\partial t} + \mu^2 V_t \frac{\partial f^2}{\partial v} + \frac{1}{2}\frac{\partial^2 f^2}{\partial v^2}(\varsigma^2)^2 V_t^2 + q_{11} f^1 - q_{22} f^2 = 0 \,, \tag{8.34}$$

subject to the boundary conditions

$$f(t, \mathbf{e}_i, L) = 1 \,, \quad f(t, \mathbf{e}_i, \infty) = 0 \,, \tag{8.35}$$

and terminal condition

$$f(t_k, \mathbf{e}_i, v) = H_k = I_{\{L \geq \min_{t\in[0,t_k]} V_t\}} \,, \tag{8.36}$$

for each $i = 1, 2$.

To approximate the solution of the system of coupled PDEs (8.33)-(8.34) subject to the conditions (8.35)-(8.36), we implement the Crank-Nicolson finite difference scheme. This finite difference scheme is essentially an average of the implicit and explicit methods and converges faster than either the implicit or the explicit scheme. For a further discussion on the numerical schemes see for example Wilmott [146].

The time domain $[0, t_k]$ is divided into p subintervals of equal length Δt and the space domain $[0, V_{\max}]$ is divided into q subintervals of equal length ΔV. For each $i = 1, 2$; $j = 1, 2, \ldots, q$; and $l = 1, 2, \ldots, p$; let $f^i(l, j) := f_i(t_l, V_j)$ denote the value of the default probability at the grid point (t_l, V_j) when the Markov chain $\mathbf{x}$ is in the i^{th} state. The Crank-Nicolson scheme leads to the following approximations to the system of coupled PDEs in the region $\{\underline{m}(t) > L\}$:

$$\begin{aligned}
&-\breve{\alpha}^1(j) f^1(l-1, j-1) + (\breve{\beta}^1(j) - 1) f^1(l-1, j) \\
&\quad + \breve{\gamma}^1(j) f^1(l-1, j+1) + \breve{d}^2(j) f^2(l-1, j) \\
&= \breve{\alpha}^1(j) f^1(l, j-1) + (-1 - \breve{\beta}^1(j)) f^1(lj) - \breve{\gamma}^1(j) f^1(l, j+1) - \breve{d}^2(j) f^2(lj) \,,
\end{aligned} \tag{8.37}$$

$$\begin{aligned}
&-\breve{\alpha}^2(j) f^2(l-1, j-1) + (\breve{\beta}^2(j-1)) f^2(l-1, j) \\
&\quad + \breve{\gamma}^2(j) f^2(l-1, j+1) + \breve{d}^1(j) f^1(l-1, j) \\
&= \breve{\alpha}^2(j) f^2(l, j-1) + (-1 - \breve{\beta}^2(j)) f^2(lj) - \breve{\gamma}^2(j) f^2(l, j+1) - \breve{d}^1(j) f^1(lj) \,,
\end{aligned} \tag{8.38}$$

where

$$\breve{\alpha}^i(j) = \left(\frac{\mu^i j}{4} - \frac{(\varsigma^i)^2 j^2}{4}\right)\Delta t, \quad \breve{\gamma}^i(j) = \left(\frac{\mu^i j}{4} + \frac{(\varsigma^i)^2 j^2}{4}\right)\Delta t, \quad \breve{d}^i(j) = \frac{q_{ii}}{2}\Delta t\ ,$$

and

$$\breve{\beta}^i(j) = \left(-\frac{(\varsigma^i)^2 j^2}{2} - \frac{q_{ii}}{2}\right)\Delta t,$$

for each $i = 1, 2$ and for $l = 1, \ldots, p$ and $j = 1, \ldots, q$.

We can rewrite equations (8.37)-(8.38) in matrix form $\mathbf{A_1} f(l-1) = \mathbf{A_2} f(l)$. The matrices $\mathbf{A_1}$ and $\mathbf{A_2}$ are constructed from the two matrices $\mathbf{B_i}$ and $\hat{\mathbf{B}}_\mathbf{i}$ and diagonal matrices $\mathbf{C_i}$ for every $i = 1, 2$ given below.

$$\mathbf{A_1} = \begin{bmatrix} \mathbf{B_1} & \mathbf{C_1} \\ \mathbf{C_2} & \mathbf{B_2} \end{bmatrix} \quad \mathbf{A_2} = \begin{bmatrix} \hat{\mathbf{B}}_\mathbf{1} & -\mathbf{C_1} \\ -\mathbf{C_2} & \hat{\mathbf{B}}_\mathbf{2} \end{bmatrix}$$

with

$$\mathbf{B_i} = \begin{pmatrix} b_0 & 0 & 0 & \cdots & 0 \\ -\breve{\alpha}^i(1) & (\breve{\beta}^i(1) - 1) & \breve{\gamma}^i(1) & 0 & 0 \\ 0 & -\breve{\alpha}^i(2) & (\breve{\beta}^i(2) - 1) & \breve{\gamma}^i(2) & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & -\breve{\alpha}^i(q-1) & (\breve{\beta}^i(q-1) - 1) & \breve{\gamma}^i(q-1) \\ 0 & \cdots & \cdots & 0 & b_1 \end{pmatrix}$$

and

$$\mathbf{C_i} = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & \breve{d}^i(1) & 0 & \cdots & 0 \\ 0 & 0 & \breve{d}^i(2) & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \breve{d}^i(q-1) & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix},$$

where $b_0 = b_1 = 1$.

The matrices $\hat{\mathbf{B}}_\mathbf{i}$, for each $i = 1, 2$, have the same structure as $\mathbf{B_i}$. However, for the matrix $\hat{\mathbf{B}}_\mathbf{i}$, its elements on the main diagonal are changed to $(-\breve{\beta}^i(j) - 1)$, it's lower diagonal elements become $\breve{\alpha}^i(j)$ and its upper diagonal elements are of the form $-\breve{\gamma}^i(j)$. The parameters for the boundary conditions are set to $\bar{b}_0 = 1$ and $\bar{b}_1 = 0$.

In the no-regime-switching case, the martingale condition gives the following PDE for the real-world default probability associated with the term t_k in the region $\{\underline{m}(t) > L\}$

$$\frac{\partial f}{\partial t} + \mu V_t \frac{\partial f}{\partial v} + \frac{1}{2}\varsigma^2 V_t^2 \frac{\partial^2 f}{\partial v^2} = 0 \,, \tag{8.39}$$

with boundary conditions

$$f(t, L) = 1 \,, \quad f(t, \infty) = 0 \,, \tag{8.40}$$

and terminal condition

$$f(t_k, v) = H_k = I_{\{L \geq \min_{t\in[0,t_k]} V_t\}} \,, \tag{8.41}$$

where μ is the real-world drift of the firm's value under $\mathcal{P}$.

The PDE in (8.39) subject to (8.40)-(8.41) is discretised and the Crank-Nicolson scheme applied. That gives

$$\begin{aligned} &-\breve{\alpha}(j)f(l-1, j-1) - (1 + \breve{\beta}(j))f(l-1, j) + \breve{\gamma}(j)f(l-1, j+1) \\ &= \breve{\alpha}(j)f(l, j-1) + (\breve{\beta}(j-1))f(l, j) - \breve{\gamma}(j)f(l, j+1) \quad , \end{aligned} \tag{8.42}$$

where

$$\breve{\alpha}(j) = \left(\frac{\mu j}{4} - \frac{\varsigma^2 j^2}{4}\right)\Delta t, \quad \breve{\beta}(j) = \frac{\varsigma^2 j^2}{2}\Delta t \quad \text{and} \quad \breve{\gamma}(j) = \left(\frac{\mu j}{4} + \frac{\varsigma^2 j^2}{4}\right)\Delta t \,.$$

This numerical scheme described is implemented in MATLAB. The results of the implementation will be presented and discussed in the section 8.5.

8.5 Numerical results and sensitivity analysis

In this section, the numerical results for the real-world default probabilities under $\mathcal{P}$ and the swap rates under the risk-neutral measure $\mathcal{P}^{\vartheta}$ determined by the Esscher transform are presented. A sensitivity analysis for the real-world default probabilities and the swap rates is performed, when the intensity parameters in

the **Q**-matrix and other model parameters are varied. For this implementation, a two-state Markov chain is considered, where state "1" and state "2" represent a "good" economy and a "bad" economy, respectively. The code of the MATLAB implementation can be found in appendix F.

8.5.1 Numerical results for actual default probabilities

The actual default probabilities are calculated under the real-world measure $\mathcal{P}$, where the drifts of the firm's value are μ_1 and μ_2 for the "good" economy and the "bad" economy, respectively. Suppose that $q_{12} = q_{21} = \lambda > 0$, where λ is interpreted as the intensity parameter of the transitions of the Markov chain. Then, the **Q**−matrix of the two-state Markov chain has the following form:

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{bmatrix}.$$

First, numerical results are presented for varying values of λ and for different choices of the constant default barrier level L. Solutions for the actual default probabilities are obtained by moving backwards in time. Two situations are considered, namely, one for the economy starting at the "good" state and another for the economy starting at the "bad" state. In this first numerical part, λ is set to $0.1, 0.3, 0.5, 0.7$ and 0.9 with different values for the default barrier level, $L = 110, 120, 130$ and 140. The default barrier level L is given exogenously and remains constant over time in each case.

The value of the firm at time $t = 0$ is set to $V_0 = 300$, which is higher than the default barrier level L in all cases. Hence, default has not yet occurred at time 0. In setting the grids for the numerical computations, the minimum value that can be reached by the value of the firm is set to $V_{min} = 1$ in both economic states. For a sensible implementation, V_{min} is set at a lower level than L to allow for the possibility that the value of the firm can drop below the default barrier level, and, hence there is a positive probability that default occurs. The maximum value that can be attained by the value of the firm is set to $V_{max} = 770$ in the "good" economic state and $V_{max} = 550$ in the "bad" economic state. This reflects that the value of the firm in the "bad" economy cannot grow as high as it can be in the "good"

economy. The drift of the firm's value in the "good" state μ^1 is set to be 10%. Since state 2 represents the "bad" economic state, μ^2 should be lower than μ^1, μ^2 is set to $\mu^2 = 6\%$. The values of other parameters are $\Delta t = \frac{1}{50}$, $\varsigma^1 = 0.2$, $\varsigma^2 = 0.4$, $p = T/\Delta t$ for maturities $T = 1, 2, ..., 12$ and $q = 70$. Note that $\varsigma^1 < \varsigma^2$ is consistent with the belief that volatility is higher in the "bad" economy than in the "good" economy.

For the implementation of equations (8.42)-(8.43) describing the no-regime-switching model, the model parameters are chosen according to the specification of the "good" state of the economy in the regime-switching case, namely, $\mu = \mu^1$ and $\varsigma = \varsigma^1$. To make the comparison valid, the grid is also set according to the one chosen for the regime-switching model. That is, $\Delta t = 1/50$ and $q = 70$. The values of the default barrier level L are identical to those values assumed in the regime-switching cases. The initial value of the firm V_0, the lower bound V_{min} and the upper bound V_{max} of the value of the firm are identical to the corresponding values assumed in the regime-switching case when the economy starts at the "good" state.

Case I: the impact of λ, L and T

Table 8.1 displays the actual default probabilities for varying values of the parameter λ in the $\mathbf{Q}-$matrix and the default barrier level L with maturities $T = 1, 2, \ldots, 12$. One may observe that the regime-switching model yields default probabilities that are higher than those in the no-regime-switching model. Since the regime-switching model allows the switching of the economy to the "bad" state, the higher risk of default in that state is reflected in the higher default probabilities arising from the regime-switching model. This illustrates that the introduction of the effect of macro-economic conditions via the Markovian regime-switching principle provides a possible way to explain and improve the underestimation of the real-world probability of default predicted by the Merton structural model.

The entries of the $\mathbf{Q}-$matrix have a clear effect on the actual default probabilities. Higher λ's result in higher default probabilities, if we start at the "good" economy at time $t = 0$. This effect may be attributed to a higher probability of switching to the "bad" economic regime with increasing values of λ. On the other

hand, if we start at the "bad" economic regime, the real-world default probabilities decrease as λ increases. This is due to an increasing probability of switching to the "good" regime with increasing values of λ. A similar effect is observed on the real-world default probabilities when the default barrier level L is set to a higher value. Since the firm's value is more likely to lie below a higher default barrier level, the real-world default probabilities in the no-regime-switching case as well as the regime-switching case are higher when L is higher. The real-world default probabilities increase as the time to maturity increase. This reflects that the longer the time to maturity the higher the default risk is.

Now, the interest rate μ^1, μ^2 and the volatilities ς^1 and ς^2 are varied to perform a sensitivity analysis for the real-world default probabilities. The entries in the $Q-$matrix are set to $\lambda = 0.1$ and the default barrier level to $L = 130$.

Case II: The impact of μ^i

Figure 8.1 depicts the real-world default probabilities calculated with varying values for μ^1 and μ^2. In the left graph, the value for μ^2 was set to 0.06 while the value of μ^1 ranges from 0.06 to 0.1, with an increment of 0.01. It becomes apparent that the real-world default probabilities are sensitive to the change in the real-world drift μ^1 of the firm's value. In all cases, the real-world default probabilities decline as μ^1 increases. The effect of μ^1 seems to be most significant in the regime-switching case starting at the "good" economy. The real-world default probabilities in the regime-switching cases are always higher than those in the corresponding no-regime-switching setting. The higher real-world default probabilities account for the possible "bad" state included in the regime-switching case starting at the "good" economy and that starting at the "bad" economy. The graph on the right of Figure 8.1 depicts changes in the real-world default probabilities when μ^2 is varied between 0.06 and 0.1, with an increment of 0.01, and μ^1 is set to 0.1. Here, the real-world default probabilities in the one-state model remain the same when μ^2 changes since the one-state model does not depend on μ^2. The real-world default probabilities decrease as μ^2 increases in the regime-switching case starting at the "good" economy and the one starting at the "bad" economy. This is similar to the impact of μ^1 on the real-world default probabilities in the regime-switching

Table 8.1: Real-world default probabilities for varying levels of λ and default barrier levels $L = 110, 120, 130$ and 140

Actual default probabilities obtained through the Crank-Nicholson scheme

Default barrier level	one state model	2 states initial state: "good" economy λ=0.1	2 states initial state: "bad" economy λ=0.1	2 states initial state: "good" economy λ=0.3	2 states initial state: "bad" economy λ=0.3	2 states initial state: "good" economy λ=0.5	2 states initial state: "bad" economy λ=0.5	2 states initial state: "good" economy λ=0.7	2 states initial state: "bad" economy λ=0.7	2 states initial state: "good" economy λ=0.9	2 states initial state: "bad" economy λ=0.9
L=110	0.00000	0.00131	0.03988	0.00326	0.03346	0.00456	0.02861	0.00543	0.02491	0.00599	0.02205
	0.00001	0.00822	0.10152	0.01769	0.07690	0.02215	0.06267	0.02425	0.05398	0.02525	0.04837
	0.00011	0.01820	0.14004	0.03490	0.10113	0.04087	0.08310	0.04318	0.07360	0.04420	0.06795
	0.00033	0.02899	0.16411	0.05079	0.11652	0.05714	0.09802	0.05948	0.08905	0.06059	0.08386
	0.00062	0.03952	0.17953	0.06443	0.12741	0.07076	0.10973	0.07315	0.10150	0.07439	0.09677
	0.00093	0.04924	0.18945	0.07580	0.13556	0.08201	0.11907	0.08449	0.11153	0.08584	0.10721
	0.00121	0.05787	0.19574	0.08511	0.14179	0.09124	0.12650	0.09381	0.11958	0.09526	0.11561
	0.00144	0.06534	0.19955	0.09266	0.14656	0.09875	0.13238	0.10141	0.12599	0.10294	0.12234
	0.00163	0.07166	0.20164	0.09871	0.15017	0.10480	0.13697	0.10754	0.13104	0.10914	0.12767
	0.00176	0.07688	0.20251	0.10349	0.15283	0.10962	0.14050	0.11245	0.13498	0.11410	0.13185
	0.00184	0.08113	0.20250	0.10723	0.15471	0.11342	0.14314	0.11631	0.13799	0.11801	0.13507
	0.00189	0.08450	0.20185	0.11010	0.15595	0.11635	0.14505	0.11930	0.14021	0.12104	0.13748
L=120	0.00000	0.00313	0.07867	0.00780	0.06624	0.01094	0.05700	0.01304	0.05004	0.01445	0.04474
	0.00006	0.01396	0.14924	0.03005	0.11432	0.03772	0.09465	0.04147	0.08294	0.04336	0.07554
	0.00033	0.02715	0.18565	0.05208	0.13649	0.06119	0.11449	0.06492	0.10322	0.06671	0.09666
	0.00080	0.04032	0.20610	0.07061	0.14970	0.07973	0.12864	0.08331	0.11870	0.08515	0.11304
	0.00134	0.05248	0.21795	0.08556	0.15874	0.09432	0.13949	0.09784	0.13074	0.09977	0.12578
	0.00183	0.06322	0.22465	0.09740	0.16528	0.10580	0.14786	0.10933	0.14007	0.11135	0.13564
	0.00225	0.07240	0.22811	0.10666	0.17005	0.11480	0.15426	0.11838	0.14724	0.12045	0.14325
	0.00256	0.08006	0.22945	0.11384	0.17347	0.12181	0.15906	0.12543	0.15267	0.12756	0.14905
	0.00278	0.08629	0.22940	0.11932	0.17581	0.12721	0.16257	0.13087	0.15670	0.13303	0.15339
	0.00291	0.09124	0.22840	0.12342	0.17727	0.13128	0.16502	0.13498	0.15961	0.13717	0.15656
	0.00298	0.09509	0.22674	0.12642	0.17802	0.13427	0.16660	0.13800	0.16158	0.14021	0.15876
	0.00299	0.09798	0.22464	0.12852	0.17818	0.13637	0.16749	0.14014	0.16280	0.14236	0.16018
L=130	0.00000	0.00452	0.10369	0.01127	0.08767	0.01584	0.07584	0.01892	0.06999	0.02101	0.06029
	0.00019	0.01759	0.17571	0.03795	0.13636	0.04788	0.11439	0.05288	0.10142	0.05553	0.09328
	0.00083	0.03259	0.21033	0.06270	0.15762	0.07412	0.13425	0.07907	0.12240	0.08159	0.11554
	0.00171	0.04711	0.22895	0.08283	0.17002	0.09415	0.14823	0.09884	0.13805	0.10136	0.13229
	0.00258	0.06020	0.23916	0.09865	0.17828	0.10943	0.15870	0.11398	0.14989	0.11653	0.14491
	0.00329	0.07151	0.24442	0.11087	0.18404	0.12113	0.16653	0.12559	0.15877	0.12817	0.15439
	0.00383	0.08100	0.24661	0.12020	0.18803	0.13004	0.17230	0.13447	0.16536	0.13706	0.16145
	0.00419	0.08876	0.24684	0.12721	0.19068	0.13677	0.17642	0.14118	0.17015	0.14377	0.16662
	0.00441	0.09495	0.24578	0.13240	0.19227	0.14176	0.17923	0.14615	0.17351	0.14874	0.17030
	0.00451	0.09977	0.24387	0.13613	0.19301	0.14535	0.18096	0.14974	0.17572	0.15233	0.17278
	0.00452	0.10341	0.24141	0.13869	0.19305	0.14783	0.18189	0.15220	0.17702	0.15479	0.17431
	0.00446	0.10606	0.23856	0.14034	0.19255	0.14940	0.18211	0.15377	0.17758	0.15635	0.17506
L=140	0.00002	0.00628	0.13209	0.01570	0.11223	0.02210	0.09765	0.02645	0.08682	0.02943	0.07867
	0.00055	0.02180	0.20328	0.04701	0.15989	0.05954	0.13587	0.06604	0.12181	0.06962	0.11307
	0.00184	0.03874	0.23539	0.07442	0.17971	0.08842	0.15529	0.09474	0.14298	0.09810	0.13593
	0.00330	0.05462	0.25185	0.09601	0.19097	0.10968	0.16870	0.11559	0.15840	0.11885	0.15261
	0.00454	0.06855	0.26023	0.11251	0.19820	0.12541	0.17848	0.13103	0.16969	0.13423	0.16477
	0.00547	0.08034	0.26393	0.12493	0.20299	0.13709	0.18554	0.14251	0.17788	0.14566	0.17359
	0.00608	0.09002	0.26477	0.13414	0.20606	0.14571	0.19049	0.15100	0.18370	0.15410	0.17990
	0.00644	0.09778	0.26382	0.14086	0.20782	0.15199	0.19379	0.15717	0.18769	0.16023	0.18428
	0.00660	0.10384	0.26172	0.14563	0.20856	0.15643	0.19579	0.16154	0.19025	0.16456	0.18716
	0.00661	0.10843	0.25889	0.14887	0.20850	0.15944	0.19676	0.16449	0.19168	0.16747	0.18886
	0.00651	0.11180	0.25559	0.15093	0.20779	0.16132	0.19691	0.16631	0.19222	0.16926	0.18962
	0.00633	0.11414	0.25198	0.15204	0.20656	0.16229	0.19641	0.16723	0.19205	0.17015	0.18965

cases. However, the impact of μ^2 on the real-world default probabilities in the regime-switching case starting at the "bad" economy is more significant than the impact of μ^1. Hence, there is an asymmetric effect of the drift on the real-world default probabilities when the economy starts at the "good" state and when it starts at the "bad" state.

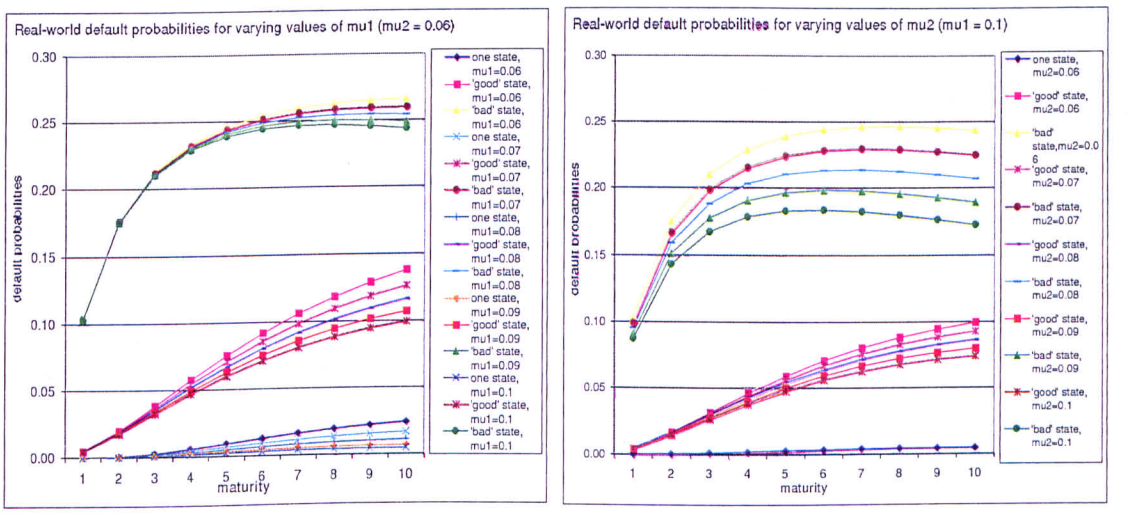


Figure 8.1: Real-world default probabilities for varying values of μ^1 and μ^2

Case III: The impact of ς^i

In Figure 8.2, a sensitivity analysis for varying values of ς^i is shown. The left graph of Figure 8.2 depicts the real-world default probabilities when ς^1 is varied between 0.1 and 0.5 with a constant value for $\varsigma^2 = 0.5$. The real-world drifts are again set to $\mu^1 = 0.1$ and $\mu^2 = 0.06$. Changing the value of ς^1 has more significant effect on the real-world default probabilities if we start at the "bad" state. In all of the three cases, namely, the one-state case, the regime-switching cases starting at the "good" state and starting at the "bad" state, the real-world default probabilities increase as ς^1 increases. The right graph depicts the real-world default probabilities when ς^2 ranges from 0.1 to 0.5 with a constant value for $\varsigma^1 = 0.1$. Here, the real-world default probabilities from the one-state model remain the same as ς^2 changes since the one-state model does not depend on the parameters in state 2. The real-world default probabilities increase as ς^2 increases in the regime-switching case starting at the "good" state and that starting at the "bad" state.

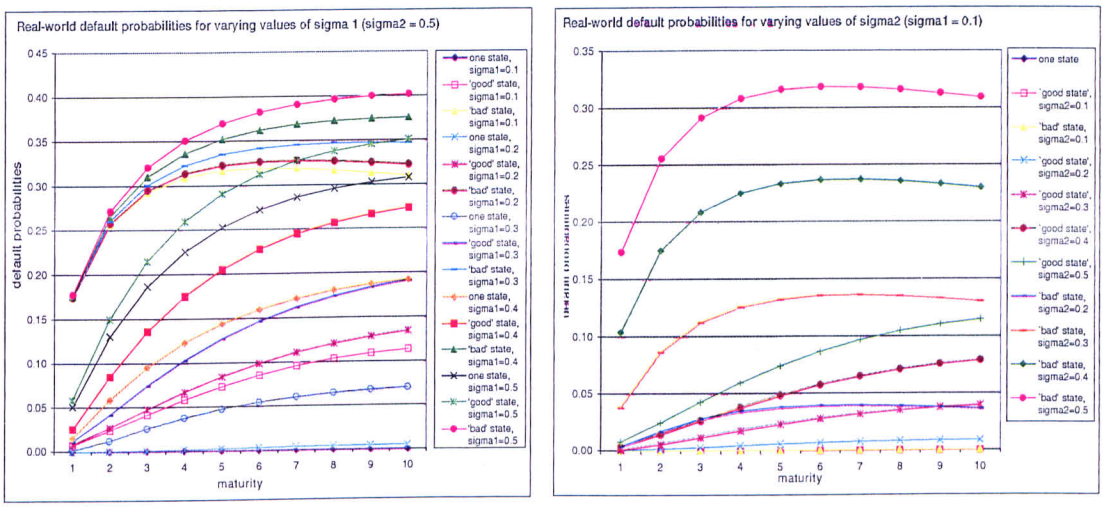


Figure 8.2: Real-world default probabilities for varying values of ς^1 and ς^2

8.5.2 Numerical results for swap rates

The swap rates of the CDS under the derived risk-neutral measure $\mathcal{P}^{\vartheta}$ are calculated. Risk-neutral default probabilities are used to calculate the swap rates. In this case, the drift of the firm's asset value is identical to the interest rate for each economic state. For this calculation, the interest rate in the "good" economic state $\breve{r}^1$ is set according to the LIBOR rate and the interest rate in the "bad" economic state is chosen to lie 1.5% lower than $\breve{r}^1$. Note that $\breve{r}^1$ and $\breve{r}^2$ are chosen to be lower than the drifts under the real-world probability measure $\mathcal{P}$. All other parameters are the same as those under the real-world setting described above.

Case I: The impact of λ, L and T

Table 8.2 depicts the swap rates calculated with the risk-neutral default probabilities for maturities $1, \ldots, 10$. The recovery rate RR is chosen to be 10% and the time duration between payments is assumed to be one year. The swap rates implied by the regime-switching model are higher than those obtained from the no-regime-switching case. This reveals that higher risk premiums are required to compensate

for higher default risk due to the possibility of switching to the "bad" economy in the regime-switching model. An assumption of being in the "good" state at time $t = 0$ in the regime-switching model leads to lower swap rates compared to the assumption of being in the "bad" state at time $t = 0$. In other words, a lower risk premium is required if the economy starts at the "good" state initially.

Here, it is clear that the intensity parameter λ in the $\mathbf{Q}-$matrix has significant effect on the swap rates. If the economy starts at the "good" state, the swap rates increase as λ increases. This may be attributed to a higher risk premium that is required to compensate for the higher risk of switching to the "bad" economic regime when λ increases. On the other hand, if we start at the "bad" economic regime, the swap rates decrease as λ increases. As λ increases, it is more likely that the state of the economy will switch from the "bad" state to the "good" state later. Since there is a lower probability of default in the "good" state, the switching from the "bad" state to the "good" state results in a reduction of risk. This, in turns, reduces the risk premium to compensate for the default risk, and hence, a lower swap rate. The swap rates increase when the default barrier level L is set to be a higher value in the no-regime-switching case as well as the regime-switching cases. Since the firm's asset value is more likely to lie below a higher default barrier level, there is higher risk that the firm will default. So, a higher risk premium is required to compensate for this higher risk. The swap rates increase as the time to maturity increases. This reflects that the longer the time to maturity the higher the risk premium is required to compensate for the higher default risk.

Case II: The impact of $\breve{r}^i$

Again, the entries in the $\mathbf{Q}-$matrix are set to $\lambda = 0.1$ and the default barrier level is set to $L = 130$. First, suppose that $\breve{r}^1$ takes values from 0.01 to 0.05, with an increment of 0.01 whilst $\breve{r}^2 = 0.01$. This ensures that the interest rate in the "good" economy is higher than that in the "bad" one. The other model parameters are the same as those chosen above.

The left graph in Figure 8.3 depicts the plots of the swap rates with different values of $\breve{r}^1$ and maturities $T = 1, 2, \ldots, 10$ from the two-state model and the one-state

Table 8.2: Swap rates for varying levels of λ and default barrier levels $L = 110, 120, 130$ and 140

Credit default swap rates calculated with risk-neutral default probabilities obtained through the Crank-Nicholson scheme

Default barrier level	one state model	2 states initial state: "good" economy λ=0.1	2 states initial state: "bad" economy λ=0.1	2 states initial state: "good" economy λ=0.3	2 states initial state: "bad" economy λ=0.3	2 states initial state: "good" economy λ=0.5	2 states initial state: "bad" economy λ=0.5	2 states initial state: "good" economy λ=0.7	2 states initial state: "bad" economy λ=0.7	2 states initial state: "good" economy λ=0.9	2 states initial state: "bad" economy λ=0.9
L=110	0.00009	0.00762	0.06599	0.01606	0.04649	0.01974	0.03687	0.02130	0.03195	0.02194	0.02933
	0.00078	0.01990	0.11940	0.03805	0.08084	0.04443	0.06608	0.04700	0.05965	0.04839	0.05636
	0.00251	0.03501	0.16301	0.06203	0.10984	0.07027	0.09333	0.07321	0.08744	0.07382	0.08595
	0.00520	0.05180	0.19987	0.08651	0.13647	0.09620	0.11977	0.10001	0.11413	0.10330	0.11057
	0.00853	0.06925	0.23105	0.11040	0.16111	0.12133	0.14492	0.12564	0.13999	0.12632	0.14014
	0.01227	0.08727	0.25974	0.13419	0.18539	0.14631	0.16981	0.15135	0.16526	0.15613	0.16177
	0.01628	0.10587	0.28770	0.15821	0.21007	0.17157	0.19501	0.17720	0.19091	0.17818	0.19218
	0.02034	0.12415	0.31353	0.18138	0.23377	0.19595	0.21920	0.20218	0.21546	0.20794	0.21237
	0.02438	0.14207	0.33798	0.20383	0.25676	0.21957	0.24262	0.22635	0.23925	0.22789	0.24116
	0.02826	0.15906	0.36028	0.22491	0.27824	0.24173	0.26451	0.24902	0.26146	0.25531	0.25901
L=120	0.00032	0.01198	0.07919	0.02496	0.05380	0.03042	0.04274	0.03260	0.03797	0.03341	0.03600
	0.00207	0.02868	0.13546	0.05427	0.08906	0.06292	0.07414	0.06620	0.06915	0.06796	0.06739
	0.00554	0.04801	0.17848	0.08420	0.11810	0.09476	0.10314	0.09822	0.09986	0.09868	0.10086
	0.01021	0.06868	0.21343	0.11359	0.14464	0.12551	0.13099	0.12986	0.12872	0.13371	0.12768
	0.01547	0.08952	0.24212	0.14145	0.16907	0.15450	0.15711	0.15924	0.15616	0.15954	0.15974
	0.02102	0.11062	0.26818	0.16868	0.19302	0.18282	0.18263	0.18825	0.18256	0.19357	0.18228
	0.02673	0.13207	0.29355	0.19584	0.21725	0.21116	0.20821	0.21709	0.20908	0.21757	0.21444
	0.03230	0.15282	0.31680	0.22169	0.24032	0.23814	0.23248	0.24462	0.23413	0.25078	0.23490
	0.03771	0.17291	0.33874	0.24646	0.26252	0.26403	0.25573	0.27097	0.25815	0.27192	0.26472
	0.04279	0.19170	0.35860	0.26943	0.28303	0.28802	0.27720	0.29540	0.28030	0.30190	0.28233
L=130	0.00098	0.01459	0.08347	0.03020	0.05639	0.03678	0.04541	0.03941	0.04125	0.04042	0.03995
	0.00476	0.03413	0.14031	0.06405	0.09254	0.07431	0.07874	0.07825	0.07514	0.08039	0.07454
	0.01089	0.05627	0.18283	0.09796	0.12222	0.11032	0.10934	0.11437	0.10806	0.11486	0.11078
	0.01820	0.07954	0.21687	0.13076	0.14920	0.14452	0.13837	0.14944	0.13850	0.15374	0.13916
	0.02582	0.10267	0.24442	0.16144	0.17379	0.17626	0.16524	0.18144	0.16703	0.18153	0.17288
	0.03347	0.12581	0.26925	0.19109	0.19769	0.20693	0.19122	0.21271	0.19418	0.21839	0.19599
	0.04108	0.14914	0.29333	0.22046	0.22171	0.23737	0.21704	0.24355	0.22123	0.24365	0.22925
	0.04833	0.17153	0.31525	0.24816	0.24439	0.26610	0.24132	0.27273	0.24655	0.27909	0.24981
	0.05523	0.19304	0.33582	0.27452	0.26604	0.29346	0.26439	0.30046	0.27063	0.30092	0.28019
	0.06160	0.21304	0.35430	0.29876	0.28588	0.31861	0.28551	0.32595	0.29265	0.33245	0.29752
L=140	0.00255	0.01761	0.08571	0.03588	0.05755	0.04358	0.04711	0.04665	0.04385	0.04782	0.04343
	0.00971	0.04043	0.14187	0.07460	0.09386	0.08638	0.08181	0.09087	0.07998	0.09328	0.08078
	0.01946	0.06569	0.18296	0.11262	0.12356	0.12656	0.11343	0.13097	0.11454	0.13130	0.11920
	0.03001	0.09171	0.21531	0.14881	0.15034	0.16406	0.14305	0.16922	0.14597	0.17372	0.14860
	0.04030	0.11714	0.24105	0.18217	0.17450	0.19830	0.17009	0.20352	0.17502	0.20309	0.18334
	0.05024	0.14228	0.26401	0.21408	0.19777	0.23102	0.19594	0.23669	0.20233	0.24236	0.20652
	0.05991	0.16743	0.28615	0.24543	0.22097	0.26326	0.22141	0.26915	0.22933	0.26848	0.24021
	0.06894	0.19136	0.30614	0.27474	0.24267	0.29340	0.24512	0.29960	0.25435	0.30569	0.26039
	0.07741	0.21420	0.32477	0.30244	0.26321	0.32191	0.26746	0.32832	0.27795	0.32783	0.29070
	0.08513	0.23527	0.34136	0.32770	0.28187	0.34790	0.28772	0.35451	0.29932	0.36048	0.30734

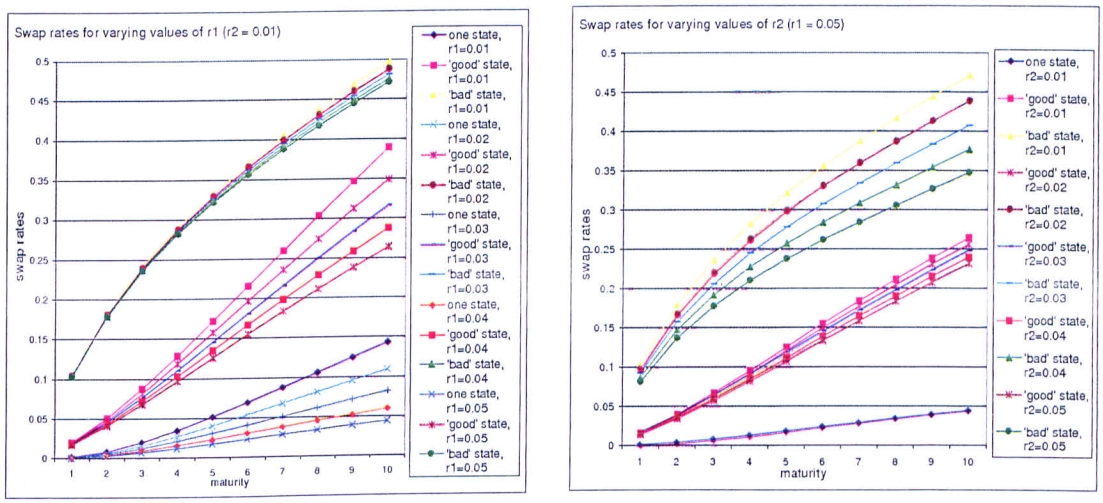


Figure 8.3: Swap rates for varying values of $\breve{r}^1$ and $\breve{r}^2$

model. With $\breve{r}^2$ being fixed, the swap rates decrease as $\breve{r}^1$ increases for both the no-regime-switching and the regime-switching cases. Since the interest rate is the expected rate of growth of the firm value and the swap rates are evaluated under the risk-neutral measure $\mathcal{P}^{\vartheta}$, an increase in the interest rate leads to a higher expected firm value, and this decreases the risk-neutral default probabilities. Therefore the swap rates decrease as the risk-neutral default probabilities decrease. From this numerical work it can be seen, that a regime-switching model that assumes a starting "bad" economic state produces the highest swap rates, therefore a higher risk premium is required in that case. In the right graph of Figure 8.3 the effect of the market interest rate $\breve{r}^2$ in the "bad" economy on the swap rates is investigated. Suppose that $\breve{r}^2$ takes values from 0.01 to 0.05, with an increment of 0.01 whilst $\breve{r}^1 = 0.05$. This ensures that $\breve{r}^2 \leq \breve{r}^1$. The other model parameters are the same as those in section 8.5.1. Independent of the state where the model is assumed to start at, increasing values of $\breve{r}^2$ lead to decreasing swap rates. The explanation of the effect of $\breve{r}^2$ on the swap rates is the same as that of $\breve{r}^1$ on the swap rates. However, $\breve{r}^2$ seems to have more significant effect on the swap rates starting in the "bad" economy than $\breve{r}^1$. This reveals that there might be an asymmetric effect between the interest rates $\breve{r}^1$ and $\breve{r}^2$ on the swap rates. The no-regime-switching case is not affected by changing the values of $\breve{r}^2$ since the values for this model depend on the parameters of the "good" economy.

Case III: The impact of ς^i

First the impact of the volatility ς^1 in the "good" economy on the swap rates is studied. Suppose that ς^1 takes values from 0.1 to 0.5, with an increment of 0.1 whilst $\varsigma^2 = 0.5$. The other model parameters are the same as those in subsection 8.5.1. The interest rate $\breve{r}^1$ is set to 0.05 and $\breve{r}^2$ is set to 0.01, lower than $\breve{r}^1$. The left graph in Figure 8.4 depicts the plots of the swap rates with different values of ς^1 for maturities $T = 1, 2, \ldots, 10$ from the two-state model and the one-state model. Here, the swap rates increase as ς^1 increases in the one-state case, the regime-switching cases starting at the "good" state and starting at the "bad" state. This reflects that as the volatility ς^1 of the firm's asset value becomes higher, a higher risk premium is required to compensate for the higher default risk induced by the higher volatility of the firm's asset value.

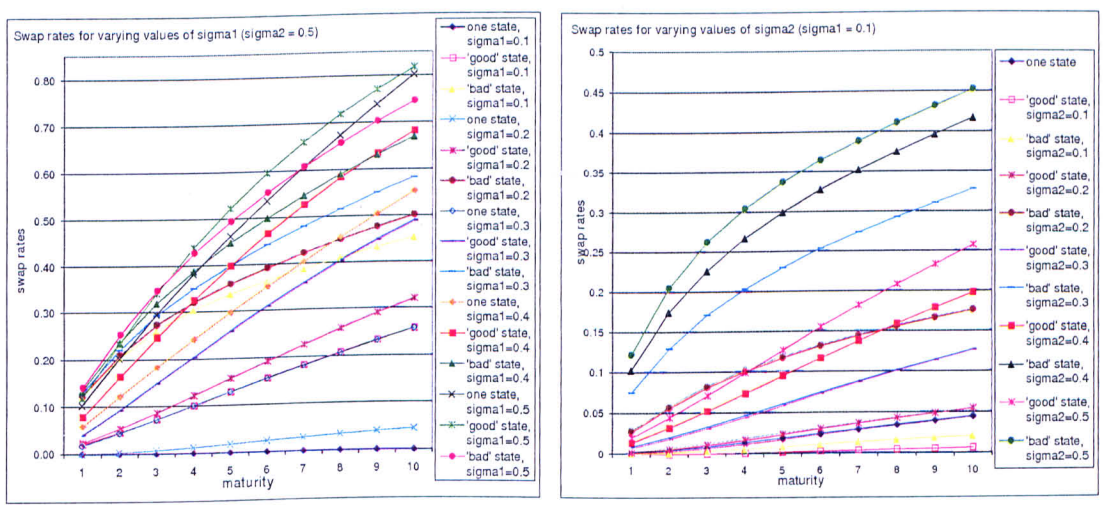


Figure 8.4: Swap rates for varying values of ς^1 and ς^2

To investigate the effect of the volatility ς^2 in the "bad" economy on the swap rates suppose that ς^2 takes values from 0.1 to 0.5, with an increment of 0.1 whilst $\varsigma^1 = 0.1$. The other model parameters are the same as those in subsection 8.5.1. The swap rates with different values of ς^2 and maturities $T = 1, 2, \ldots, 10$ from the two-state model and the one-state model are plotted in the right graph of Figure 8.4. The no-regime-switching model is not affected by varying the values for ς^2.

The swap rates increase as the volatility ς^2 of the firm's asset value increases for the regime-switching cases starting at the "good" economy and starting at the "bad" economy.

8.6 Monte-Carlo simulation

The default probabilities and swap rates obtained from the numerical scheme in the previous section are now compared to results from a Monte-Carlo simulation for the regime-switching case. The parameters are chosen to be identical to those used within the numerical scheme. Again, actual default probabilities are calculated for varying default barrier levels L and intensities λ. For each default probability 10,000 paths for the firm value are generated. The results are depicted in Table 8.3. The default probabilities in the non-regime-switching case are calculated with the analytical formula from section 8.2, equation (8.21).

In addition, swap rates were calculated with risk-neutral default probabilities obtained via a Monte-Carlo simulation for the regime-switching case. The risk-neutral default probabilities are simulated with the same parameters as in the previous section; 10,000 paths are generated for each default probability. The default barrier L is set to 130 and $\lambda = 0.1$. The starting value of the firm is $V_0 = 300$. Table 8.4 shows the calculated swap rates for the non-regime-switching and the regime-switching case, respectively. As in the previous Table, the swap rates from the non-regime-switching case are based on default probabilities obtained through the analytical formula.

The results for both the actual default probabilities as well as the swap rates are roughly close to the results from the PDE technique. However, the Crank-Nicolson scheme is computational more efficient than the Monte-Carlo simulation. The average CPU-time of simulating default probabilities for one barrier level for the five different values of λ with the Monte-Carlo technique is 83 seconds, while the Crank-Nicolson scheme leads to these results in about 5.5 seconds. There is therefore a large computational benefit in applying the PDE technique.

Table 8.3: Actual default probabilities for varying levels of λ and default barrier levels $L = 110, 120, 130$ and 140 using a Monte-Carlo simulation

Actual default probabilities obtained through Monte-Carlo simulation in the 2-state case and the analytical formula for the 1-state case

Default barrier level	one state model analytical solution	2 states initial state: "good" economy λ=0.1	2 states initial state: "bad" economy λ=0.1	2 states initial state: "good" economy λ=0.3	2 states initial state: "bad" economy λ=0.3	2 states initial state: "good" economy λ=0.5	2 states initial state: "bad" economy λ=0.5	2 states initial state: "good" economy λ=0.7	2 states initial state: "bad" economy λ=0.7	2 states initial state: "good" economy λ=0.9	2 states initial state: "bad" economy λ=0.9
L=110	0.00000	0.00000	0.00670	0.00000	0.00820	0.00490	0.00340	0.00760	0.00000	0.00630	0.00000
	0.00005	0.00010	0.04700	0.00000	0.05240	0.01730	0.01610	0.01220	0.00750	0.01090	0.00950
	0.00042	0.00020	0.10360	0.00000	0.10670	0.03320	0.03700	0.04820	0.01370	0.04460	0.01540
	0.00126	0.00060	0.15830	0.00190	0.15210	0.05370	0.05530	0.06010	0.03930	0.05320	0.04410
	0.00245	0.00090	0.19790	0.00740	0.18670	0.08200	0.07530	0.08820	0.05730	0.09700	0.05220
	0.00381	0.00330	0.24930	0.01570	0.21900	0.10200	0.10060	0.11020	0.08340	0.10240	0.08830
	0.00520	0.00290	0.30580	0.02690	0.24780	0.11000	0.12280	0.12930	0.09550	0.13280	0.08840
	0.00654	0.00340	0.32180	0.04040	0.25550	0.14330	0.14240	0.14460	0.11830	0.14020	0.12440
	0.00779	0.00480	0.34890	0.05640	0.27460	0.15540	0.15680	0.16530	0.13100	0.15960	0.13020
	0.00894	0.00470	0.39320	0.06120	0.28870	0.16510	0.17030	0.17920	0.15770	0.17560	0.15370
	0.00997	0.00570	0.42410	0.07850	0.30340	0.18660	0.18690	0.19450	0.16630	0.19470	0.16160
	0.01089	0.00760	0.43920	0.09280	0.31350	0.19850	0.19520	0.21060	0.17400	0.20780	0.18730
L=120	0.00000	0.00000	0.01290	0.00000	0.01100	0.00480	0.00630	0.01250	0.00000	0.01180	0.00000
	0.00017	0.00010	0.06530	0.00030	0.06610	0.02280	0.02480	0.01820	0.01390	0.01570	0.01600
	0.00107	0.00060	0.12620	0.00070	0.12860	0.04800	0.05060	0.06290	0.02180	0.06340	0.02080
	0.00274	0.00080	0.19080	0.00300	0.18900	0.07250	0.07510	0.08000	0.05660	0.06850	0.06030
	0.00480	0.00160	0.24090	0.00990	0.22090	0.10390	0.09150	0.11370	0.07680	0.11710	0.06560
	0.00696	0.00340	0.28490	0.02250	0.24890	0.11630	0.12300	0.13420	0.10200	0.12270	0.10280
	0.00905	0.00500	0.32820	0.03480	0.27960	0.14360	0.13940	0.16530	0.12850	0.16010	0.11650
	0.01097	0.00690	0.36150	0.04930	0.29630	0.16590	0.16590	0.17800	0.14780	0.16450	0.14640
	0.01271	0.00580	0.40000	0.06690	0.31160	0.18390	0.17720	0.20020	0.16390	0.20180	0.15360
	0.01426	0.00890	0.41240	0.08030	0.32950	0.19860	0.19400	0.20830	0.17400	0.20030	0.17830
	0.01563	0.00910	0.44830	0.09430	0.33360	0.20260	0.21400	0.23120	0.19680	0.22890	0.18670
	0.01683	0.01120	0.47820	0.11550	0.34600	0.23180	0.23100	0.24170	0.21310	0.23110	0.20380
L=130	0.00001	0.00000	0.02170	0.00000	0.02280	0.01160	0.01120	0.02290	0.00000	0.02280	0.00000
	0.00051	0.00010	0.08760	0.00020	0.08450	0.03410	0.03450	0.02960	0.02170	0.02960	0.02320
	0.00245	0.00150	0.16550	0.00090	0.16050	0.06680	0.06880	0.08490	0.03260	0.08850	0.02860
	0.00539	0.00320	0.21360	0.00590	0.21740	0.09410	0.10200	0.09670	0.07250	0.09680	0.07950
	0.00866	0.00340	0.27690	0.01370	0.25110	0.12560	0.12740	0.14460	0.09270	0.14040	0.08660
	0.01184	0.00610	0.32560	0.02770	0.27870	0.14300	0.15120	0.16140	0.12750	0.15300	0.13410
	0.01477	0.00780	0.35960	0.04720	0.30710	0.16780	0.17170	0.19640	0.14580	0.18810	0.14020
	0.01738	0.00940	0.39670	0.06090	0.33050	0.19080	0.19700	0.19890	0.16830	0.20450	0.18010
	0.01967	0.01260	0.42860	0.08090	0.34010	0.20590	0.20890	0.22510	0.19900	0.23290	0.18790
	0.02166	0.01290	0.46000	0.09670	0.36560	0.23250	0.22400	0.24080	0.20400	0.23530	0.21090
	0.02339	0.01170	0.48740	0.11300	0.37940	0.23860	0.24820	0.25700	0.22440	0.25950	0.21730
	0.02489	0.01560	0.50270	0.12770	0.38130	0.25840	0.25190	0.26880	0.24220	0.26250	0.23950
L=140	0.00003	0.00000	0.03530	0.00000	0.03150	0.01680	0.01640	0.03170	0.00000	0.03220	0.00010
	0.00135	0.00070	0.11350	0.00060	0.11060	0.04290	0.05080	0.04880	0.03280	0.04110	0.03200
	0.00503	0.00220	0.19440	0.00220	0.19220	0.08530	0.07910	0.10930	0.04200	0.11300	0.04600
	0.00979	0.00480	0.25690	0.00650	0.25490	0.11950	0.11640	0.12950	0.09460	0.12490	0.09770
	0.01459	0.00780	0.30830	0.01880	0.29960	0.15000	0.14810	0.16410	0.12080	0.17900	0.11470
	0.01901	0.00910	0.36270	0.03890	0.32650	0.18410	0.17430	0.19620	0.14420	0.18390	0.15850
	0.02289	0.01300	0.38990	0.05820	0.34200	0.20150	0.20190	0.21650	0.17730	0.22910	0.16640
	0.02625	0.01290	0.43870	0.07300	0.36480	0.22110	0.21720	0.24310	0.20660	0.23730	0.19930
	0.02914	0.01570	0.45990	0.09550	0.37050	0.24360	0.23400	0.25650	0.21950	0.26540	0.21220
	0.03160	0.01550	0.48600	0.11290	0.39490	0.26450	0.24950	0.27090	0.23500	0.27730	0.24620
	0.03371	0.01740	0.51540	0.13480	0.41170	0.27050	0.28050	0.29610	0.25690	0.29820	0.23990
	0.03551	0.02100	0.52700	0.14660	0.41440	0.28580	0.28450	0.30040	0.26600	0.29990	0.26360

Table 8.4: Swap rates for varying levels of λ and default barrier levels $L = 110, 120, 130$ and 140 using a Monte-Carlo simulation

Credit default swap rates calculated with risk-neutral default probabilities obtained through Monte-Carlo simulation in the 2-state case and the analytical formula for the 1-state case

Default barrier level	one state model analytical default probabilities	2 states initial state: "good" economy λ=0.1	2 states initial state: "bad" economy λ=0.1	2 states initial state: "good" economy λ=0.3	2 states initial state: "bad" economy λ=0.3	2 states initial state: "good" economy λ=0.5	2 states initial state: "bad" economy λ=0.5	2 states initial state: "good" economy λ=0.7	2 states initial state: "bad" economy λ=0.7	2 states initial state: "good" economy λ=0.9	2 states initial state: "bad" economy λ=0.9
L=110	0.00024	0.00126	0.03569	0.00533	0.02513	0.01502	0.01283	0.01528	0.00787	0.00748	0.01114
	0.00229	0.00468	0.09646	0.01927	0.06751	0.03735	0.03938	0.04219	0.02641	0.04703	0.02220
	0.00726	0.01161	0.15860	0.04397	0.10536	0.06729	0.07137	0.07368	0.05326	0.06691	0.05647
	0.01493	0.02247	0.22187	0.06751	0.14279	0.10656	0.09872	0.11032	0.09181	0.11298	0.08464
	0.02524	0.03888	0.29523	0.09886	0.19657	0.14124	0.13286	0.15191	0.12265	0.15337	0.12158
	0.03703	0.05701	0.35376	0.13196	0.24359	0.18149	0.17589	0.20065	0.16496	0.19613	0.16027
	0.04944	0.07945	0.42505	0.17238	0.29700	0.22292	0.21971	0.24916	0.20565	0.23819	0.20557
	0.06313	0.10337	0.50451	0.20638	0.34299	0.27508	0.28061	0.29544	0.24288	0.29473	0.24517
	0.07790	0.13445	0.57566	0.24849	0.39523	0.32395	0.31174	0.34181	0.28161	0.34992	0.28816
	0.09414	0.16021	0.64993	0.30168	0.44188	0.37275	0.37314	0.39442	0.33311	0.39460	0.34476
L=120	0.00080	0.00162	0.04951	0.00796	0.04149	0.01668	0.01981	0.02174	0.01499	0.01242	0.01617
	0.00545	0.00642	0.11776	0.02408	0.08766	0.04932	0.04828	0.05821	0.03686	0.05928	0.03185
	0.01456	0.01854	0.20035	0.05705	0.13130	0.08449	0.08497	0.10017	0.07088	0.08159	0.07963
	0.02706	0.03405	0.27940	0.08680	0.18053	0.12844	0.12903	0.14592	0.10939	0.15253	0.10170
	0.04271	0.05225	0.34903	0.12225	0.23510	0.17535	0.17713	0.19429	0.15913	0.16874	0.15737
	0.05976	0.07046	0.41397	0.17250	0.28436	0.21814	0.21497	0.24604	0.20114	0.23443	0.20309
	0.07717	0.10093	0.50856	0.21325	0.33653	0.27625	0.27429	0.29455	0.25381	0.27975	0.25707
	0.09597	0.11867	0.58050	0.25457	0.40733	0.31926	0.32863	0.35559	0.30486	0.34671	0.30682
	0.11593	0.16181	0.66307	0.30102	0.46601	0.37230	0.39253	0.40280	0.35009	0.38515	0.35144
	0.13756	0.19126	0.74678	0.34887	0.53197	0.44143	0.43346	0.46250	0.41083	0.45936	0.41762
L=130	0.00225	0.00234	0.06724	0.01198	0.04146	0.03112	0.02818	0.03164	0.01951	0.01811	0.02408
	0.01149	0.01111	0.15145	0.03666	0.09338	0.06745	0.06335	0.07288	0.05203	0.07635	0.04589
	0.02667	0.02597	0.24255	0.07232	0.16088	0.10669	0.10671	0.11878	0.09761	0.10952	0.09936
	0.04561	0.04330	0.32604	0.11155	0.21174	0.16624	0.15629	0.17723	0.14436	0.18222	0.13746
	0.06799	0.06626	0.41035	0.14875	0.26235	0.22044	0.21667	0.23589	0.19722	0.21481	0.19439
	0.09150	0.09338	0.49552	0.20325	0.33053	0.25958	0.27086	0.29586	0.25839	0.28581	0.25267
	0.11497	0.12092	0.59185	0.25807	0.41033	0.32952	0.32277	0.35515	0.30453	0.33545	0.31450
	0.13989	0.15816	0.67550	0.30764	0.48221	0.38359	0.38937	0.40271	0.36151	0.41087	0.38746
	0.16606	0.19079	0.77591	0.37020	0.53778	0.45271	0.45396	0.49140	0.42621	0.46568	0.43264
	0.19413	0.22423	0.85170	0.41871	0.60714	0.52455	0.50591	0.54656	0.48894	0.53634	0.49452
L=140	0.00545	0.00442	0.07877	0.01860	0.05574	0.03443	0.03521	0.04161	0.02967	0.02174	0.03497
	0.02199	0.01564	0.18636	0.04739	0.11905	0.08241	0.08427	0.09157	0.06604	0.08738	0.06710
	0.04540	0.03429	0.27648	0.09012	0.18781	0.13538	0.13620	0.14987	0.12436	0.13593	0.13207
	0.07247	0.05586	0.36016	0.13202	0.24951	0.19209	0.19688	0.21077	0.17149	0.20619	0.18359
	0.10306	0.08361	0.46202	0.18312	0.31172	0.25406	0.24948	0.26379	0.23285	0.26595	0.24199
	0.13431	0.11992	0.56059	0.24372	0.37245	0.30765	0.32094	0.33632	0.29561	0.34731	0.31481
	0.16501	0.15602	0.69785	0.31016	0.45877	0.38551	0.37909	0.41249	0.37125	0.40490	0.37743
	0.19721	0.19404	0.76204	0.35820	0.54127	0.46204	0.45862	0.47302	0.44724	0.47955	0.44623
	0.23076	0.22752	0.86368	0.43799	0.62902	0.53069	0.52873	0.54977	0.51332	0.54540	0.51668
	0.26643	0.27420	0.96514	0.49026	0.69787	0.60149	0.58931	0.63005	0.57154	0.61764	0.58568

8.7 Some concluding remarks

The pricing of a CDS under a Markov-modulated version of Merton's structural model gives the parameters the possibility to switch between economic regimes. In this pricing model, the continuous-time, finite-state Markov chain is supposed to be observable. An equivalent martingale measure in this incomplete market setting can be determined through the Esscher transform. The Crank-Nicolson scheme is adopted to provide a numerical solution for a system of coupled PDE's for both the risk-neutral and the real-world default probabilities. The results indicate that in a regime-switching model the presence of a "bad" economic state increases the swap rate of a CDS substantially and therefore an empirically observed underestimation of default probabilities through the Merton model can be improved through adding possible regime shifts. In the sensitivity analysis of the numerical solutions, we found that increasing the value of λ increases the real-world default probabilities and the swap rates assuming we start at the "good" economy and a decreasing effect otherwise. A significant effect on the numerical approximation through changes in parameter values is observed, quantifying the parameters is therefore an important issue for evaluating the default probabilities. Finally a comparison of the Crank-Nicolson scheme and a Monte-Carlo method gives evidence of the significantly higher computational efficiency of the numerical scheme.

Chapter 9

Conclusions and directions for future work

9.1 Summary of contributions

In this thesis, regime-switching models for various financial variables such as commodity prices, interest rates, electricity spot prices, stock indices and firm's value processes were developed. The shift between regimes is governed by a hidden Markov process and this flexibility of the models allowing different parameter values for different regimes captures the dynamic changes occurring in the market economy. Throughout the entire course of estimating parameters, the change of probability measure technique coupled with the EM-algorithm was applied. We provided refined, extended and new recursive optimal parameter estimation procedures via the calculated adaptive filters both in the univariate and multivariate cases. Each regime-switching model was tested on a financial data set. This gives important insights on the performance of the model in forecasting the value or price levels of a particular financial variable and the optimal regime-switching model that can describe the market data being analysed.

Additionally, this thesis made contributions to each of the three main practical goals of financial modelling: (i) In the first part of chapter 7 trading strategies and a mean-variance portfolio selection criterion were developed, which presented a certain perspective in performing optimisation under an HMM setting; (ii) The

latter part of chapter 7 combines aspects of optimisation and risk measurement by including CVaR constraints in asset allocation problems; and (iii) chapter 8 demonstrated how to price a particular derivative under a regime-switching framework, in which case the market is no longer complete. In addition, chapter 6 contains a section on the computation of an expected spot on delivery that illustrates an application of the filtering results to pricing.

Other major highlights of research endeavours pursued in this thesis include the following:

- The predictability of the daily gold spot prices using an HMM-based model was investigated. On the basis of the Diebold-Kilian metric, the 2-state and 3-state HMMs yield higher predictability measures than those of the ARCH(1) and GARCH(1,1) models.

- Based on a general filtering technique described in chapter 4, optimal parameter estimates are derived for an OU-type model and an OU-type model with added jump component. The pure OU-type model was implemented on Canadian T-bill yields data whilst the OU with jump model was implemented on Nord pool data of daily electricity spot prices. In particular, the model for electricity spot prices is assumed to follow an exponential OU-process plus a jump term and this exponential is scaled by a deterministic sinusoidal function that takes into account the seasonality component of electricity prices. Both models have parameters that shift regimes and therefore possess the ability to capture the observable features of the market data. These models modulated by HMM are self-tuning as parameter values are updated the moment new information is obtained from the market. The important empirically observed characteristics of the short term interest rates and the electricity market prices are captured by the models as evidenced by low forecast errors. Similar trends are exhibited by the forecasts that closely resemble the dynamics and patterns of the actual time series data.

- Asset allocation problems under a Markov-switching market were addressed. A scenario generator was developed to stochastic programming problems in a CVaR setting. Based on historical data, optimal parameter estimates are

calculated which are then employed for the generation of scenarios. The generated scenarios proved to lead to stable solutions within the portfolio optimisation problems considered in chapter 7.

9.2 Future directions

Several research questions naturally arise as an immediate consequence of the results and analyses of this work. These possible research enquiries are outlined below.

- In the study of the gold market in chapter 3, the question of the optimal number of states in the HMM model setting still needs to be addressed more fully in later works. Recent approaches that could be applied can be found in the papers of Olteanu [124], and Strikholm and Teräsvirta [141]. It would also be worth exploring the statistical procedures in Hansen ([82] and [83]), and formal hypothesis tests and other related statistical procedures on the determination of states for those HMM-based models developed in this thesis.

- The filtering algorithms in chapter 4 were successfully implemented in chapters 5 and 6. The forecasts we obtained are judged reasonable by the error analyses performed in those chapters. However, these models can be used further as frameworks in the valuation of financial instruments. The model in forecasting short term interest rates can be utilised for pricing bonds when an equivalent martingale measure is constructed, perhaps through the Esscher transform methodology discussed in chapter 8. The pricing of ESD within the HMM set-up for electricity process can be extended to longer maturities.

- A possible alternative formulation of the electricity spot price model could be carried out further.[1] For example, a seasonal function can be chosen by fitting the deterministic function to the data set without jumps first and then

[1] I am grateful for the helpful suggestions of John Braun and Matt Davison concerning the proposed electricity spot price model in this thesis. Comments of other attendees in the 06 December 2007 colloquium held in the Department of Statisitcal and Actuarial Sciences at the University of Western Ontario, Canada, where parts of this work were presented, are also acknowledged.

adding an indicator function for jumps. Also, when the data series contains both weekday and weekend prices, there is the so-called on/off peak issue. It is known a priori that weekends are going to be different than weekdays as electricity demand is usually lower. It is postulated that perhaps the weekly dependence can be better modelled by fitting a parametric function that contains an indicator-type function, that is, assign 1 for weekdays and 0 for weekends and holidays.

- For the asset allocation problem, we wish to extend and test the scenario generator for multiple time step optimisation problems.

- The model proposed in pricing CDS in chapter 8 needs to be capable of calibration to make this usable for practitioners. This should be compared with popular credit risk models like the Expected Default Frequency^{TM} (EDF^{TM}) credit measure supplied by the KMV corporation, the underlying principles of which are also based on Merton's theory. The EDF^{TM} credit measure represents the probability of default in the forthcoming year. A further possibility in this area is the pricing of complex credit products like multi-name credit derivatives where correlation has to be included considering that this is already a multi-dimensional problem.

- The proposed HMMs and the on-line estimation schemes here can be certainly adopted and applied to other financial data sets. It would be worthwhile to analyse high frequency data where appropriate dynamics for the observation process are chosen within the HMM setting. The implementation and perhaps extension or modification of filtering algorithms developed in the thesis designed to efficiently handle data that arrive in minutes or seconds need to be examined further.

Appendices

Appendix A

Appendix for Chapter 2

The Bayes theorem together with its proof is re-visited here as this theorem is central to the optimal estimation of parameters in this thesis.

Definition A.1

The probability measure P is absolutely continuous with respect to the probability measure $\widetilde{P}$, written $P \ll \widetilde{P}$, if $P(A) = 0$ implies $\widetilde{P}(A) = 0$ for each $A \in \mathcal{F}$. If $P \ll \widetilde{P}$ and also $\widetilde{P} \ll P$, the two measures are said to be equivalent and write $P \equiv \widetilde{P}$.

Theorem A.2 (Conditional Bayes' theorem)

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra and suppose ϖ is any $\mathcal{G}$-measurable function. Assume further that $\widetilde{P}$ is a probability measure equivalent to P via the Radon-Nikodým derivative

$$\frac{d\widetilde{P}}{dP} = \Lambda \, .$$

Then

$$\widetilde{E}[\varpi \mid \mathcal{G}] = \frac{E[\Lambda \varpi \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]} \, .$$

Proof

To prove that this equation holds, we have to show that

$$\int_A \widetilde{E}[\varpi \mid \mathcal{G}] \, d\widetilde{P} = \int_A \frac{E[\Lambda \varpi \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]} \, d\widetilde{P} \;\; \forall A \in \mathcal{G} \, .$$

We define a measurable function ψ

$$\psi = \begin{cases} E[\Lambda\varpi \mid \mathcal{G}] & \text{if } E[\Lambda \mid \mathcal{G}] > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

Now we have to distinguish between two subsets of $\mathcal{G}$. *Suppose* $G = \{\omega : E[\Lambda \mid \mathcal{G}] = 0\}$ *and* $G^c = \{\omega : E[\Lambda \mid \mathcal{G}] > 0\}$. *So,* $\Lambda = 0$ *a.s. on* G. *Let* A *be any set in* $\mathcal{G}$ *Then* $A = B \cup C$ *with* $B = A \cap G^c$ *and* $C = A \cap G$.

With this distinction we have

$$\begin{aligned} \int_A \widetilde{E}[\varpi \mid \mathcal{G}]\, d\widetilde{P} &= \int_A \varpi\, d\widetilde{P} \\ &= \int_A \varpi\Lambda\, dP = \int_B \varpi\Lambda\, dP + \underbrace{\int_C \varpi\Lambda\, dP}_{=0}. \end{aligned} \tag{A.2}$$

From the first integral in (A.2) *we get*

$$\int_B \Lambda\varpi\, dP = E[I_B\varpi\Lambda]. \tag{A.3}$$

Now using (A.1)

$$\begin{aligned} \int_B \psi\, d\widetilde{P} = & \int_B \frac{E[\Lambda\varpi \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]}\, d\widetilde{P} \\ = & \widetilde{E}\left[I_B \frac{E[\Lambda\varpi \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]}\right] = E\left[I_B\Lambda \frac{E[\Lambda\varpi \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]}\right]. \end{aligned} \tag{A.4}$$

We apply the tower property to (A.4) *and obtain*

$$\begin{aligned} E\left[I_B\Lambda \frac{E[\Lambda\varpi \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]}\right] &= E\Big[E\big[I_B\Lambda \frac{E[\Lambda\varpi \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]} \mid \mathcal{G}\big]\Big] \\ &= E\left[I_B E[\Lambda \mid \mathcal{G}] \frac{E[\Lambda\varpi \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]}\right] \\ &= E[I_B\Lambda\varpi]. \end{aligned} \tag{A.5}$$

Hence from equation (A.3) *and* (A.5) $\int_B \Lambda\varpi\, dP = \int_B \psi\, d\widetilde{P}$. *using* (A.2) *we therefore see that*

$$\begin{aligned} \int_A \Lambda\varpi\, dP &= \int_C \Lambda\varpi\, dP + \int_B \Lambda\varpi\, dP \\ &= \int_A \widetilde{E}[\varpi \mid \mathcal{G}]\, d\widetilde{P} = \int_A \psi\, d\widetilde{P}. \end{aligned}$$

So we finally get

$$\widetilde{E}[\varpi \mid \mathcal{G}] = \frac{E[\Lambda\varpi \mid \mathcal{G}]}{E[\Lambda \mid \mathcal{G}]} \,.$$

□

Appendix B

Appendix for Chapter 3

B.1 Proof of theorem 3.2

Proof

From the definition of $\eta_l(J_l^{(sr)}X_l)$ we have

$$
\begin{aligned}
\eta_l(J_l^{(sr)}\mathbf{x}_l) &= \bar{\mathrm{E}}\left[\Lambda_l J_l^{(sr)}\mathbf{x}_l|\mathcal{F}_l^y\right] \\
&= \bar{\mathrm{E}}\left[\Lambda_{l-1}\lambda_l\left(J_{l-1}^{(sr)} + \langle\mathbf{x}_{l-1},\mathbf{e}_r\rangle\langle\mathbf{x}_l,\mathbf{e}_s\rangle\right)\mathbf{x}_l|\mathcal{F}_l^y\right] \\
&= \bar{\mathrm{E}}\left[\Lambda_{l-1}\lambda_l J_{l-1}^{(sr)}\left(\mathbf{\Pi}\mathbf{x}_{l-1}+\mathbf{v}_l\right)|\mathcal{F}_l^y\right] + \bar{\mathrm{E}}\left[\Lambda_{l-1}\langle\mathbf{x}_{l-1},\mathbf{e}_r\rangle|\mathcal{F}_l^y\right]\frac{\phi(\sigma_r^{-1}(y_l-f_r))}{\sigma_r\phi(y_l)}\pi_{sr}\mathbf{e}_s \\
&= \sum_{i=1}^{N}\bar{\mathrm{E}}\left[\Lambda_{l-1}\langle\mathbf{x}_{l-1},\mathbf{e}_i\rangle J_{l-1}^{(sr)}|\mathcal{F}_l^y\right]\frac{\phi(\sigma_i^{-1}(y_l-f_i))}{\sigma_i\phi(y_l)}\mathbf{\Pi}\mathbf{e}_i \\
&\qquad + \bar{\mathrm{E}}\left[\Lambda_{l-1}\langle\mathbf{x}_{l-1},\mathbf{e}_r\rangle|\mathcal{F}_l^y\right]\frac{\phi(\sigma_r^{-1}(y_l-f_r))}{\sigma_r\phi(y_l)}\pi_{sr}\mathbf{e}_s \\
&= \sum_{i=1}^{N}\left\langle\eta_{l-1}(J_{l-1}^{(sr)}\mathbf{x}_{l-1}),\mathbf{e}_i\right\rangle\frac{\phi(\sigma_i^{-1}(y_l-f_i))}{\sigma_i\phi(y_l)}\mathbf{\Pi}\mathbf{e}_i + \langle\Xi_{l-1},\mathbf{e}_r\rangle\frac{\phi(\sigma_r^{-1}(y_l-f_r))}{\sigma_r\phi(y_l)}\pi_{sr}\mathbf{e}_s \\
&= \mathbf{\Pi}\boldsymbol{D}_l(y_l)\eta_{l-1}(J_{l-1}^{(sr)}\mathbf{x}_{l-1}) + \langle\Xi_{l-1},\mathbf{e}_r\rangle\frac{\phi(\sigma_r^{-1}(y_l-f_r))}{\sigma_r\phi(y_l)}\pi_{sr}\mathbf{e}_s \ .
\end{aligned}
$$

The proofs of two other recursive formulae follow similar arguments above by using the definition of λ_l in equation (3.4) and evaluating the resulting conditional expectation under $\bar{P}$.

□

B.2 Proof of theorem 3.3

Proof for formula (3.11)*:*

Using the Radon-Nikodým derivative of $P^{\widehat{\theta}}$ with respect to P^{θ} in (3.10) we have

$$\begin{aligned}
\log \frac{dP^{\widehat{\theta}}}{dP^{\theta}} &= \sum_{l=1}^{k} \log \left(\sum_{s,r=1}^{N} \left(\frac{\widehat{\pi}_{sr}}{\pi_{sr}} \right)^{\langle \mathbf{x}_l, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle} \right) \\
&= \sum_{l=1}^{k} \sum_{s,r=1}^{N} (\log \widehat{\pi}_{sr} - \log \pi_{sr}) \langle \mathbf{x}_l, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \\
&= \sum_{s,r=1}^{N} J_k^{(sr)} \log \widehat{\pi}_{sr} + \text{Remainder} \qquad \text{(B.1)}
\end{aligned}$$

where the Remainder does not involve $\widehat{\pi}_{sr}$.

Observe that $\sum_{s=1}^{N} J_k^{(sr)} = O_k^{(r)}$ hence

$$\sum_{s=1}^{N} \widehat{J}_k^{(sr)} = \widehat{O}_k^{(r)}. \qquad \text{(B.2)}$$

The $\widehat{\pi}_{ji}$'s optimal estimate is the value that maximises the log-likelihood (B.1) subject to the constraint $\sum_{s=1}^{N} \widehat{\pi}_{sr} = 1$.

Now, introducing the Lagrange multiplier β we consider the function

$$L(\widehat{\pi}, \beta) = \sum_{r,s=1}^{N} \widehat{J}_k^{sr} \log \widehat{\pi}_{sr} + \beta \left(\sum_{s=1}^{N} \widehat{\pi}_{sr} - 1 \right) + \text{Remainder}. \qquad \text{(B.3)}$$

Differentiating (B.3) with respect to $\widehat{\pi}_{ji}$ and β and equating the derivatives to 0, we have

$$\frac{1}{\widehat{\pi}_{ji}} \widehat{J}_k^{(ji)} + \beta = 0 \qquad \text{(B.4)}$$

and

$$\sum_{s=1}^{N} \widehat{\pi}_{sr} = 1. \qquad \text{(B.5)}$$

Equation (B.4) can be re-written as

$$\widehat{\pi}_{ji} = \frac{\widehat{J}_k^{(ji)}}{-\beta}. \qquad \text{(B.6)}$$

Consequently,

$$\sum_{j=1}^{N} \widehat{\pi}_{ji} = \frac{\sum_{j=1}^{N} \widehat{J}_k^{(ji)}}{-\beta}. \tag{B.7}$$

But from (B.5) and (B.2), equation (B.7) simplifies to

$$1 = \frac{\widehat{O}_k^{(i)}}{-\beta} \quad \text{or} \quad -\beta = \widehat{O}_k^{(i)}.$$

Hence, from equation (B.6), the optimal estimate for $\widehat{\pi}_{ji}$ is

$$\widehat{\pi}_{ji} = \frac{\widehat{J}_k^{(ji)}}{\widehat{O}_k^{(i)}} = \frac{\eta_k(J_k^{(ji)})}{\eta_k(O_k^{(i)})},$$

which is the desired result in (3.11). In terms of the recursive filters involving the vectors $J\,\mathbf{x}$ and $O\,\mathbf{x}$,

$$\widehat{\pi}_{ji} = \frac{\left\langle \eta_k(J_k^{(ji)}\mathbf{x}_k), \mathbf{1} \right\rangle}{\left\langle \eta_k(O_k^{(i)}\mathbf{x}_k), \mathbf{1} \right\rangle}.$$

□

Proof of formula (3.12)*:*

Let $\mathbf{f} = (f_1, f_2, \ldots, f_N)^\top \in I\!R^N$. To perform a change to $\widehat{\mathbf{f}} = (\widehat{f}_1, \widehat{f}_2, \ldots, \widehat{f}_N)^\top \in I\!R^N$, consider a new measure $P^{\widehat{\theta}}$ defined by

$$\left.\frac{dP^{\widehat{\theta}}}{dP^{\theta}}\right|_{\mathcal{F}_k} = \Lambda_k^* = \prod_{l=1}^{k} \lambda_l^*$$

where

$$\lambda_l^* = \exp\left(\frac{1}{2\left\langle \boldsymbol{\sigma}, \mathbf{x}_{l-1}\right\rangle^2}\left(\left\langle \mathbf{f}, \mathbf{x}_{l-1}\right\rangle^2 - \left\langle \widehat{\mathbf{f}}, \mathbf{x}_{l-1}\right\rangle^2 - 2y_l\left\langle \mathbf{f}, \mathbf{x}_{l-1}\right\rangle + 2y_l\left\langle \widehat{\mathbf{f}}, \mathbf{x}_{l-1}\right\rangle\right)\right).$$

This means that

$$\begin{aligned}
L(\widehat{\mathbf{f}}) &= E\left[\log\left.\frac{dP^{\widehat{\theta}}}{dP^{\theta}}\right| \mathcal{F}_k^y\right] \\
&= E\left[\sum_{l=1}^{k} \frac{1}{2\left\langle \boldsymbol{\sigma}, \mathbf{x}_{l-1}\right\rangle^2}\left\{\left\langle \mathbf{f}, \mathbf{x}_{l-1}\right\rangle^2 - \left\langle \widehat{\mathbf{f}}, \mathbf{x}_{l-1}\right\rangle^2 - 2y_l\left\langle \mathbf{f}, \mathbf{x}_{l-1}\right\rangle + 2y_l\left\langle \widehat{\mathbf{f}}, \mathbf{x}_{l-1}\right\rangle\right\}\right| \mathcal{F}_k^y\right] \\
&= E\left[\sum_{l=1}^{k}\sum_{r=1}^{N} \left\langle \mathbf{x}_{l-1}, \mathbf{e}_r\right\rangle \frac{f_r^2 - \widehat{f}_r^2 - 2y_l(f_r - \widehat{f}_r)}{2\sigma_r^2}\right| \mathcal{F}_k^y\right] \\
&= E\left[\sum_{r=1}^{N} \frac{2T_k^{(r)}(y)\widehat{f}_r - O_k^{(r)}\widehat{f}_r^2}{2\sigma_r^2} + \text{Remainder}\right| \mathcal{F}_k^y\right]
\end{aligned}$$

where the Remainder does not contain $\widehat{\mathbf{f}}$. Thus,

$$L(\widehat{\mathbf{f}}) = \sum_{r=1}^{N} \frac{2\widehat{T}_k^{(r)}(y)\widehat{f}_r - \widehat{O}_k^{(r)}\widehat{f}_r^2}{2\sigma_r^2} + \text{Remainder}. \tag{B.8}$$

We differentiate (B.8) in $\widehat{f}_i$ and equate the resulting derivative to 0. This gives the optimal choice for $\widehat{f}_i$ given the observations $y_1, y_2, \ldots, y_k$ and therefore

$$\widehat{f}_i = \frac{\widehat{T}_k^{(i)}(y)}{\widehat{O}_k^{(i)}} = \frac{\eta_k(T^{(i)}(y))}{\eta_k(O_k^{(i)})},$$

which is what we wanted to show in (3.12).

□

Proof of formula (3.13)*:*

Suppose we are given the vector of parameters $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)^\top \in I\!R^N$ and we wish to change to the parameters $\widehat{\boldsymbol{\sigma}} = (\widehat{\sigma}_1, \widehat{\sigma}_2, \ldots, \widehat{\sigma}_N)^\top \in I\!R^N$. Then, we define the Radon-Nikodým derivative of $P^{\widehat{\theta}}$ with respect to P^θ as $\frac{dP^{\widehat{\theta}}}{dP^\theta}\Big|_{\mathcal{F}_k} = \Lambda_k^{**}$ where $\Lambda_k^{**} = \prod_{l=1}^k \lambda_l^{**}$ and

$$\lambda^{**} = \langle \boldsymbol{\sigma}, \mathbf{x}_l \rangle \langle \widehat{\boldsymbol{\sigma}}, \mathbf{x}_l \rangle \exp\left(\frac{1}{2\langle \widehat{\boldsymbol{\sigma}}, \mathbf{x}_l \rangle^2} (y_{l+1} - \langle \mathbf{f}, \mathbf{x}_l \rangle)^2 - \frac{1}{2\langle \boldsymbol{\sigma}, \mathbf{x}_l \rangle^2} (y_{l+1} - \langle \mathbf{f}, \mathbf{x}_l \rangle)^2 \right).$$

Thus,

$$\log \frac{dP^{\widehat{\theta}}}{dP^\theta} = \sum_{l=1}^{k} \left[-\log \langle \mathbf{x}_{l-1}, \widehat{\boldsymbol{\sigma}} \rangle - \frac{1}{2\langle \widehat{\boldsymbol{\sigma}}, \mathbf{x}_{l-1} \rangle^2} (y_l - \langle \mathbf{f}, \mathbf{x}_{l-1} \rangle)^2 + \text{Remainder} \right]$$

where the Remainder is free of the term $\widehat{\sigma}$.

Hence,

$$\begin{aligned}
&E\left[\log \frac{dP^{\widehat{\theta}}}{dP} \middle| \mathcal{F}_k^y \right] \\
&= E\left[\sum_{l=1}^{k} \sum_{r=1}^{N} \left(-\langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \log \widehat{\sigma}_r - \frac{\langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle}{2\widehat{\sigma}_r^2} (y_l^2 - 2f_r y_l + f_r^2) \middle| \mathcal{F}_k^y \right) \right] \\
&\quad + \text{Remainder} \\
&= \sum_{r=1}^{N} \left[\log \widehat{\sigma}_r \widehat{O}_k^{(r)} + \frac{1}{2\widehat{\sigma}_r^2} (\widehat{T}_k^r(y^2) - 2f_r \widehat{T}_k^{(r)}(y) + f_r^2 \widehat{O}_k^{(r)}) \right] + \text{Remainder}
\end{aligned}$$

We differentiate the above expression and set its derivative to 0. Solving for the optimal choice of $\widehat{\sigma}_i^2$ given the information in $\mathcal{Y}_k$, we have

$$\widehat{\sigma}_i^2 = \frac{\eta_k(T_k^{(i)}(y^2)) - 2\widehat{f}_i \eta_k(T_k^{(i)}(y)) + \widehat{f}_i^2 \eta_k(O_k^{(i)})}{\eta_k(O_k^{(i)})},$$

which agrees with equation (3.13).

□

B.3 MATLAB source code: Implementation of the 3-state HMM for gold prices

Here, the MATLAB source code for the implementation of the 3-state hidden Markov model for the daily spot gold prices is presented. The implementation for the 2- and 4-state HMM is performed with a similar code including necessary changes for the number of states.

```
%% This program calculates the parameter estimation and the goldprice
%% forecast within a 3-state Markov chain model %%
clear;
%% load daily goldprice data between 1973 and 2006 %%
load goldprice73_06
%% set program parameters %%
state=3;
forecaststep=40;
for h=1:forecaststep
    C=data;
    K=log_returns;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% initial values %%
    mu=[-0.1; 0.04; 0.1];
    sigma=[0.3; 0.1; 0.3];
    A=[1/3 1/3 1/3; 1/3 1/3 1/3; 1/3 1/3 1/3];
    xi(:,1)=[0.9 0.05 0.05]';
    E=eye(state);
%% first conditional expectation p=E[X_1|Y_1] %%
    x1=xi(:,1);
```

```
    nenner1=0;
    for i=1:state
        nenner1=nenner1+x1(i);
    end
    p1=x1./nenner1;
    xhat(:,1)=p1;
%% Lambda for calculating start values for the recursion of the gammas %%
    w=(K(1)-(mu(1)*x1(1)+mu(2)*x1(2)+mu(3)*x1(3)))/...
        (sigma(1)*x1(1)+sigma(2)*x1(2)+sigma(3)*x1(3));
    w1=((2*pi)^(-1/2))*exp(-(w^2)/2);
    w2=((2*pi)^(-1/2))*exp(-(K(1)^2)/2);
    Lambda1=w1/((sigma(1)*x1(1)+sigma(2)*x1(2)+sigma(3)*x1(3))*w2);
%% first diagonal matrix for first gammas %%
    for i=1:3
        Y1(i)=(K(2)-mu(i)*x1(i))/(sigma(i)*x1(i));
        %% standard normal distribution of Y %%
        Z1(i)=((2*pi)^(-1/2))*exp(-(Y1(i)^2)/2);
        %% standard normal distribution of K %%
        N1(i)=(sigma(i)*x1(i))*(((2*pi)^(-1/2))*exp(-(K(2)^2)/2));
        D1(i)=Z1(i)/N1(i) ;
    end
    %% diagonal matrix %%
    DD1=diag(D1);
    xi(:,2)=A*DD1*xi(:,1);
    x2=xi(:,2);
    nenner2=0;
    for i=1:state
        nenner2=nenner2+x2(i);
    end
    p2=x2./nenner2;
    xhat(:,2)=p2;
%% first gammas for the recursion formulas %%
    for i=1:state
        gammaj1(:,i)=Lambda1*p1(i)*p2(1)*p2;
        gammaj2(:,i)=Lambda1*p1(i)*p2(2)*p2;
        gammaj3(:,i)=Lambda1*p1(i)*p2(3)*p2;
    end
    for i=1:state
        gammao(:,i)=Lambda1*p1(i)*p2;
```

```
    end
    for i=1:state
        gammat(:,i)=Lambda1*p1(i)*K(1)*p2;
    end
    for i=1:state
        gammatq(:,i)=Lambda1*p1(i)*K(1)^2*p2;
    end
%% set number of data point in one algorithm pass %%
    batch=10;
    %% interval borders %%
    a=2;
    e=batch+a;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% main for-loop to calculate filters for Markov chain processes %%
    for u=1:842
    %% batches of ten data points %%
    B=K(a:e);
    n=length(B);
    %% Calculations for xi %%
    for k=1:batch
        hh=xi(:,a+k-1);
        hh1=0;
        for i=1:state
            hh1=hh1+hh(i);
        end
    %% diagonal matrix DD %%
        for i=1:state
            Y(k,i)=(B(k+1)-mu(i)*hh(i))/(sigma(i)*hh(i));
            Z(k,i)=((2*pi)^(-1/2))*exp(-(Y(k,i)^2)/2);
            N(k,i)=(sigma(i)*hh(i))*(((2*pi)^(-1/2))*exp(-(B(k+1)^2)/2));
            D(k,i)=Z(k,i)/N(k,i) ;
        end
        DD=diag(D(k,:));
        xi(:,a+k)=(A*diag(D(k,:))*xi(:,a+k-1));
        ff1=xi(:,a+k);
        xhat(:,a+k)=ff1./(ff1(1)+ff1(2)+ff1(3));
    end
%% GAMMAS %%
    for k=1:batch
```

```
        sx=xi(:,a+k-1);
        hsum=0;
        for i=1:state
            hsum=hsum+sx(i);
        end
        for r=1:state
            y(k,r)=(B(k+1)-mu(r)*sx(r))/(sigma(r)*sx(r));
            %% numerator %%
            z(k,r)=((2*pi)^(-1/2))*exp(-(y(k,r)^2)/2);
             %% denumerator %%
            n(k,r)=(sigma(r)*sx(r))*(((2*pi)^(-1/2))*exp(-(B(k+1)^2)/2));
            sxi=xi(:,a+k-1);
            hsum=sxi(1)+sxi(2)+sxi(3);
            %% Gamma(J1rX)_k %%
            gammaj1(:,r)=(A*diag(D(k,:))*(gammaj1(:,r))+sxi(r)*(z(k,r)/...
                n(k,r))*A(1,r)*E(:,1));
            %% Gamma(J2rX)_k %%
            gammaj2(:,r)=(A*diag(D(k,:))*(gammaj2(:,r))+sxi(r)*(z(k,r)/...
                n(k,r))*A(2,r)*E(:,2));
            %% Gamma(J3rX)_k %%
            gammaj3(:,r)=(A*diag(D(k,:))*(gammaj3(:,r))+sxi(r)*(z(k,r)/...
                n(k,r))*A(3,r)*E(:,3));
            ADXI=A*diag(D(k,:))*xi(:,a+k-1);
            %% Gamma(DrX)_k %%
            gammao(:,r)=(A*diag(D(k,:))*gammao(:,r)+sxi(r)*(z(k,r)/...
                n(k,r))*A*E(:,r));
            %% Gamma(Tr(y)X)_k %%
            gammat(:,r)=(A*diag(D(k,:))*gammat(:,r)+sxi(r)*(z(k,r)/...
                n(k,r))*B(k+1)*A*E(:,r));
            %% Gamma(Tr(y^2)X)_k
            gammatq(:,r)=(A*diag(D(k,:))*gammatq(:,r)+sxi(r)*(z(k,r)/...
            n(k,r))*B(k+1)^2*A*E(:,r));
        end
    end
%%%%%%%%%%% forecast %%%%%%%%%%%%%%%
    for k=1:batch
        chi=xi(:,a+k-2);
        b=0;
        for i=1:state
```

```
            b=b+chi(i);
        end
        %% p_hat_k=E[X_k |Y_k] %%
        phat=chi./b ;
        d=(A^h-1)*phat;
        sum1=d(1)*exp(mu(1)+0.5*sigma(1)^2)+d(2)*exp(mu(2)+0.5*sigma(2)^2);
        S(k+h)=C(a+k-2)*sum1;
    end
    F(:,u)=S(1+h:10+h)';
%%%%%%%%%%%%%% parameter updates %%%%%%%%%%%%%%%%%%%%
    gammaj1z=[0 0 0];
    for i=1:state
        gammaj11(:,i)=gammaj1(:,i);
        for j=1:state
            gammaj1z(i)=gammaj1z(i)+gammaj11(j,i);
        end
    end
    gammaj2z=[0 0 0];
    for i=1:state
        gammaj22(:,i)=gammaj2(:,i);
        for j=1:state
            gammaj2z(i)=gammaj2z(i)+gammaj22(j,i);
        end
    end
    gammaj3z=[0 0 0];
    for i=1:state
        gammaj33(:,i)=gammaj3(:,i);
        for j=1:state
            gammaj3z(i)=gammaj3z(i)+gammaj33(j,i);
        end
    end
    gammaos=[0 0 0];
    for i=1:state
        gammao1(:,i)=gammao(:,i);
        for j=1:state
            gammaos(i)=gammaos(i)+gammao1(j,i);
        end
    end
    gammats=[0 0 0];
```

```
for i=1:state
    gammat1(:,i)=gammat(:,i);
    for j=1:state
        gammats(i)=gammats(i)+gammat1(j,i);
    end
end
gammatqs=[0 0 0];
for i=1:state
    gammatq1(:,i)=gammatq(:,i);
    for j=1:state
        gammatqs(i)=gammatqs(i)+gammatq1(j,i);
    end
end
%% transition probabilities %%
for i=1:state
    AA(1,i)=gammaj1z(i)/gammaos(i);
    AA(2,i)=gammaj2z(i)/gammaos(i);
    AA(3,i)=gammaj3z(i)/gammaos(i);
end
A=AA;
%% sigma %%
for i=1:state
    sigma(i)=sqrt((gammatqs(i)-2*mu(i)*gammats(i)+...
                    mu(i)^2*gammaos(i))/gammaos(i));
end
%% mu %%
for i=1:state
    mu(i)=gammats(i)/gammaos(i);
end
M1(u,:)=mu(1);
M2(u,:)=mu(2);
M3(u,:)=mu(3);
SI1(u,:)=sigma(1);
SI2(u,:)=sigma(2);
SI3(u,:)=sigma(3);
PROB1(u,:)=A(1,1);
PROB2(u,:)=A(1,2);
PROB3(u,:)=A(1,3);
PROB4(u,:)=A(2,1);
```

```
    PROB5(u,:)=A(2,2);
    PROB6(u,:)=A(2,3);
    PROB7(u,:)=A(3,1);
    PROB8(u,:)=A(3,2);
    PROB9(u,:)=A(3,3);
    %% next interval %%
    a=a+batch;
    e=e+batch;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% plot forecast and actual data %%
forecast1(:,h)=F(:); m=length(forecast1(:,h)); t=h+1;
forecast(1:t-1,h)=zeros(t-1,1); forecast(t:m,h)=forecast1(1:m-h,h);
K=(1:m); datap=data(1:m); plot(K,forecast(:,h), K, datap)
Forecast3state=forecast save Forecast3state Forecast3state
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% loss-functions for h-step ahead forecast %%
quadratic1=(datap(t:m)-forecast(t:m,h)).^2;
quadratic(h)=mean(quadratic1);
absmeandev1=abs(datap(t:m)-forecast(t:m,h));
absmeandev(h)=mean(absmeandev1);
fourthdegree1=(datap(t:m)-forecast(t:m,h)).^4;
fourthdegree(h)=mean(fourthdegree1);
end
%% Diebold-Kilian metric for different loss-functions %%
%% quadratic loss function %%
for i=1:10
    DKquadratic10(i)=1-quadratic(i)/quadratic(10);
end
for i=1:20
    DKquadratic20(i)=1-quadratic(i)/quadratic(20);
    DKquadratic30(i)=1-quadratic(i)/quadratic(30);
    DKquadratic40(i)=1-quadratic(i)/quadratic(40);
end
%% mean absolute deviation %%
for i=1:10
    DKabsmeandev10(i)=1-absmeandev(i)/absmeandev(10);
end
for i=1:20
```

```
    DKabsmeandev20(i)=1-absmeandev(i)/absmeandev(20);
    DKabsmeandev30(i)=1-absmeandev(i)/absmeandev(30);
    DKabsmeandev40(i)=1-absmeandev(i)/absmeandev(40);
end
%% quartic loss function %%
for i=1:10
    DKfourthdegree10(i)=1-fourthdegree(i)/fourthdegree(10);
end
for i=1:20
    DKfourthdegree20(i)=1-fourthdegree(i)/fourthdegree(20);
    DKfourthdegree30(i)=1-fourthdegree(i)/fourthdegree(30);
    DKfourthdegree40(i)=1-fourthdegree(i)/fourthdegree(40);
end
-----------------------------------------------
```

Appendix C

Appendix for Chapter 5

C.1 Optimal estimate for α

To derive an optimal estimate for α, consider a new measure $\hat{P}$, which is defined by

$$\left.\frac{d\hat{P}}{dP}\right|_{\mathcal{F}_k^y} := \Lambda_k^\alpha = \prod_{l=1}^{k} \lambda_l^\alpha$$

where

$$\begin{aligned}
\lambda_l^\alpha &= \frac{\exp\left(-\frac{1}{2\xi^2(\mathbf{x}_l)}(y_{l+1}^2 + (\hat{\alpha}(\mathbf{x}_l)y_l)^2 + \gamma(\mathbf{x}_l)^2 - 2y_{l+1}\hat{\alpha}(\mathbf{x}_l)y_l - 2y_{l+1}\gamma(\mathbf{x}_l) + 2\hat{\alpha}(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l))\right)}{\exp\left(-\frac{1}{2\xi^2(\mathbf{x}_l)}(y_{l+1}^2 + (\alpha(\mathbf{x}_l)y_l)^2 + \gamma(\mathbf{x}_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l - 2y_{l+1}\gamma(\mathbf{x}_l) + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l))\right)} \\
&= \exp\Big(\frac{1}{2\xi^2(\mathbf{x}_l)}(\alpha(\mathbf{x}_l)y_l)^2 - (\hat{\alpha}(\mathbf{x}_l)y_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l \\
&\qquad + 2y_{l+1}\hat{\alpha}(\mathbf{x}_l)y_l + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l) - 2\hat{\alpha}(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l)\Big).
\end{aligned}$$

This means that

$$\begin{aligned}
\log\Lambda_k^\alpha &= \sum_{l=1}^{k}\big[(\alpha(\mathbf{x}_l)y_l)^2 - (\hat{\alpha}(\mathbf{x}_l)y_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l + 2y_{l+1}\hat{\alpha}(\mathbf{x}_l)y_l \\
&\qquad + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l) - 2\hat{\alpha}(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l)\big]/2\xi^2(\mathbf{x}_l) \\
&= \sum_{l=1}^{k}\Big(\sum_{i=1}^{n}\langle\mathbf{x}_l, e_i\rangle\big(\alpha_i^2y_l^2 - \hat{\alpha}_i^2y_l^2 - 2y_{l+1}y_l\alpha_i + 2y_{l+1}y_l\hat{\alpha}_i + 2y_l\alpha_i\gamma_i - 2\hat{\alpha}_iy_l\gamma_i\big)/2\xi_i^2\Big) \\
&= \sum_{l=1}^{k}\Big(\sum_{i=1}^{n}\langle\mathbf{x}_l, e_i\rangle\big(-\hat{\alpha}_i^2y_l^2 + 2y_{l+1}y_l\hat{\alpha}_i - 2\hat{\alpha}_iy_l\gamma_i\big)/2\xi_i^2\Big) + R(\alpha_i) \\
&= \sum_{i=1}^{n}\big(-T_k^i(y^2)\hat{\alpha}_i^2 + 2T_k^i(y_{k+1}, y_k)\hat{\alpha}_i - 2T_k^i(y)\gamma_i\hat{\alpha}_i\big)/2\xi_i^2 + R(\alpha_i) \qquad \text{(C.1)}
\end{aligned}$$

where $R(\alpha_i)$ is a remainder that does not contain $\hat{\alpha}$. The expectation of the log-likelihood conditional on $\mathcal{F}_k^y$ is

$$L(\hat{\alpha}) = E\Big[\log \Lambda_k^{\alpha} | \mathcal{F}_k^y\Big] \quad = \quad \sum_{i=1}^{n} \big(-\hat{T}_k^i(y^2)\hat{\alpha}_i^2 + 2\hat{T}_k^i(y_{k+1}, y_k)\hat{\alpha}_i - 2\hat{T}_k^i(y)\gamma_i\hat{\alpha}_i\big)/2\xi_i^2 + R(\alpha_i)$$

where $\hat{H}_k = E[H_k|\mathcal{F}_k^y]$. We differentiate $L(\hat{\alpha})$ in $\hat{\alpha}_i$ and equate the result to 0. This gives us the optimal choice of the parameter $\hat{\alpha}$. In particular,

$$(-2\hat{\alpha}_i\hat{T}_k^i(y^2) + 2\hat{T}_k^i(y_{k+1}, y_k) - 2\hat{T}_k^i(y)\gamma_i)/2\xi_i^2 = 0,$$

and solving for $\hat{\alpha}_i$, we get

$$\hat{\alpha}_i = \frac{\hat{T}_k^i(y_{k+1}, y_k) - \hat{T}_k^i(y)\gamma_i}{\hat{T}_k^i(y^2)}.$$

C.2 Optimal estimate for γ

To calculate the optimal estimate $\hat{\gamma}_i$ we consider again the following Radon-Nikodým derivative

$$\left.\frac{d\hat{P}}{dP}\right|_{\mathcal{F}_k^y} \quad := \quad \Lambda_k^{\gamma} = \prod_{l=1}^{k} \lambda_l^{\gamma}$$

with

$$\begin{aligned}
\lambda_l^{\gamma} \quad &= \quad \frac{\exp\left(-\frac{(y_{l+1}^2 + (\alpha(\mathbf{x}_l)y_l)^2 + \hat{\gamma}(\mathbf{x}_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l - 2y_{l+1}\hat{\gamma}(\mathbf{x}_l) + 2\alpha(\mathbf{x}_l)y_l\hat{\gamma}(\mathbf{x}_l))}{2\xi^2(\mathbf{x}_l)}\right)}{\exp\left(-\frac{(y_{l+1}^2 + (\alpha(\mathbf{x}_l)y_l)^2 + \gamma(\mathbf{x}_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l - 2y_{l+1}\gamma(\mathbf{x}_l) + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l))}{2\xi^2(\mathbf{x}_l)}\right)} \\
&= \quad \exp\Big(\frac{1}{2\xi^2(\mathbf{x}_l)}\big(\gamma(\mathbf{x}_l)^2 - \hat{\gamma}(\mathbf{x}_l)^2 - 2y_{l+1}\gamma(\mathbf{x}_l) + 2y_{k+1}\hat{\gamma}(\mathbf{x}_l) \\
&\qquad\qquad + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l) - 2\alpha(\mathbf{x}_l)y_l\hat{\gamma}(\mathbf{x}_l)\big)\Big) .
\end{aligned}$$

Now

$$\begin{aligned}
\log \Lambda_k^{\gamma} \quad &= \quad \sum_{l=1}^{k} \big(\gamma(\mathbf{x}_l)^2 - \hat{\gamma}(\mathbf{x}_l)^2 - 2y_{l+1}\gamma(\mathbf{x}_l) + 2y_{l+1}\hat{\gamma}(\mathbf{x}_l) \\
&\qquad\qquad + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l) - 2\alpha(\mathbf{x}_l)y_l\hat{\gamma}(\mathbf{x}_l)\big)/2\xi^2(\mathbf{x}_l) \\
&= \quad \sum_{l=1}^{k}\Big(\sum_{i=1}^{n}\langle \mathbf{x}_l, e_i\rangle\big(-\hat{\gamma}_i^2 + 2y_{k+1}\hat{\gamma}_i - 2y_k\alpha_i\hat{\gamma}_i\big)/2\xi_i\Big) + R(\gamma_i) \qquad \text{(C.2)}
\end{aligned}$$

where $R(\gamma_i)$ is independent of $\hat{\gamma}_i$. Thus

$$L(\hat{\gamma}_i) = E\Big[\log \Lambda_k^{\gamma} | \mathcal{F}_k^y\Big] = \sum_{i=1}^{n} \big(-\hat{O}_k^i \hat{\gamma}_i^2 + 2\hat{T}_{k+1}^i(y)\hat{\gamma}_i - 2\hat{T}_k^i(y)\alpha_i\hat{\gamma}_i\big)/2\xi_i + R(\gamma_i).$$

We differentiate $L(\hat{\gamma}_i)$ and set the derivative to 0. The optimal choice of $\hat{\gamma}_i$ is given by

$$\hat{\gamma}_i \;=\; \frac{\hat{T}_{k+1}^i(y) - \hat{T}_k^i(y)\alpha_i}{\hat{O}_k^i} \quad .$$

C.3 Optimal estimate for ξ

To find the optimal estimate $\hat{\xi}$, we start with the Radon-Nikodým derivative

$$\left.\frac{d\hat{P}}{dP}\right|_{\mathcal{F}_k^y} := \Lambda_k^{\xi} \;=\; \prod_{l=1}^{k} \lambda_l^{\xi}$$

where

$$\begin{aligned}
\lambda_l^{\xi} \;&=\; \frac{\frac{1}{\hat{\xi}\sqrt{2\pi}} \exp\left(-\frac{y_{l+1}^2 + (\alpha(\mathbf{x}_l)y_l)^2 + \gamma(\mathbf{x}_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l - 2y_{l+1}\gamma(\mathbf{x}_l) + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l)}{2\hat{\xi}^2(\mathbf{x}_l)}\right)}{\frac{1}{\xi\sqrt{2\pi}} \exp\left(-\frac{y_{l+1}^2 + (\alpha(\mathbf{x}_l)y_l)^2 + \gamma(\mathbf{x}_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l - 2y_{l+1}\gamma(\mathbf{x}_l) + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l)}{2\xi^2(\mathbf{x}_l)}\right)} \\
&=\; \frac{\xi(\mathbf{x}_l)}{\hat{\xi}(\mathbf{x}_l)} \exp\Big(\frac{1}{2\xi^2(\mathbf{x}_l)}\big(y_{l+1} - \alpha(\mathbf{x}_l)y_l - \gamma(\mathbf{x}_l)\big)^2 \\
&\qquad\qquad - \frac{1}{2\hat{\xi}^2(\mathbf{x}_l)}\big(y_{l+1} - \alpha(\mathbf{x}_l)y_l - \gamma(\mathbf{x}_l)\big)^2\Big) \quad .
\end{aligned}$$

Therefore the log-likelihood is

$$\log \Lambda_k^{\xi} \;=\; \sum_{l=1}^{k}\Big(- \log(\hat{\xi}(\mathbf{x}_k)) - \frac{1}{2\hat{\xi}^2(\mathbf{x}_l)}(y_{l+1} - \alpha(\mathbf{x}_l)y_l - \gamma(\mathbf{x}_l))^2\Big) + R(\xi) \tag{C.3}$$

where $R(\xi)$ is the remainder independent of $\hat{\xi}$. Hence,

$$\begin{aligned}
L(\hat{\xi}) \;&=\; E\Big[\sum_{l=1}^{k}\sum_{i=1}^{n}\Big(-\langle \mathbf{x}_l, e_i\rangle \log \hat{\xi}_i - \frac{\langle \mathbf{x}_l, e_i\rangle}{2\hat{\xi}_i^2} \\
&\qquad * \big(y_{l+1}^2 + (\alpha_i y_l)^2 + \gamma_i^2 - 2y_{l+1}\alpha_i y_l - 2y_{l+1}\gamma_i + 2\alpha_i y_l \gamma_i\big)\Big)\Big|\mathcal{F}_k^y\Big] \\
&=\; \sum_{i=1}^{n}\Big(-\log \hat{\xi}_i (\hat{O})_k^i - \frac{1}{2\hat{\xi}_i^2}\hat{T}_{k+1}^i(y^2) - \frac{\alpha_i^2}{2\hat{\xi}_i^2}\hat{T}_k^i(y^2) - \frac{\gamma_i^2}{2\hat{\xi}_i^2}\hat{O}_k^i \\
&\qquad + \frac{\alpha_i}{\hat{\xi}_i^2}\hat{T}_k^i(y_{k+1}, y_k) + \frac{\gamma_i}{\hat{\xi}_i^2}\hat{T}_{k+1}^i(y) - \frac{\alpha_i\gamma_i}{\hat{\xi}_i^2}\hat{T}_k^i(y)\Big) + R(\xi) \quad .
\end{aligned}$$

After differentiating $L(\hat{\xi})$ with respect to $\hat{\xi}$ and maximising $L(\hat{\xi})$ the optimal estimate for ξ may be shown as

$$\hat{\xi}_i = \sqrt{\frac{\hat{T}^i_{k+1}(y^2) + \alpha_i^2 \hat{T}^i_k(y^2) + \gamma_i^2 \hat{O}^i_k - 2\alpha_i \hat{T}^i_k(y_{k+1}, y_k) - 2\gamma_i \hat{T}^i_{k+1}(y) + 2\alpha_i\gamma_i \hat{T}^i_k(y)}{\hat{O}^i_k}}$$

C.4 MATLAB source code: Implementation of the 3-state HMM for interest rates

The MATLAB source code for the implementation of the 3-state hidden Markov model for interest rates is detailed below. The calculation of the AIC is included in this code. Some formulae have to be changed accordingly, when dealing with a different number of states.

```
%% Hidden Markov model for interest rates %%
%% data: daily 30-day T-bill rates 1973-2007 %%
clear;
%% read in original data for interest rate %%
load Tbill30days9605
interestrate=TBILL9607;
%% states of markov chain %%
state = 3;
%% number of data points within one batch %%
batch=20;
%% number of algorithm passes %%
no_passes=144;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% initial values for parameters %%
alpha=[0.9966+0.3 0.9966 0.6966]
```

```
gamma=[0.0078 0.0048 0.0028]
nu=[1-3*0.0105 1-6*0.0105 1-8*0.0105]
Pi=[1/3 1/3 1/3; 1/3 1/3 1/3; 1/3 1/3 1/3];
%%initial value for markov chain%%
X=[1 0 0]';
E=eye(state);
%% initial Radon-Nikodym derivative for change of measure %%
alphaIP=0;
gammaIP=0;
nuIP=0;
for i=1:state
 a1=alpha(i)*X(i);
 alphaIP=alphaIP+a1;
 b1=gamma(i)*X(i);
 gammaIP=gammaIP+b1;
 c1=nu(i)*X(i);
 nuIP=nuIP+c1;
end
z(1)=0;
Lambda1=exp(-((alphaIP*interestrate(1)+gammaIP)/(nuIP^2))*interestrate(2)...
    -1/2*((alphaIP*interestrate(1)+gammaIP)/nuIP)^2);
LogLikelihood1=-0.5*log(2*pi*nuIP^2)-(interestrate(1)-alphaIP-gammaIP)^2/...
    (2*nuIP^2);
X2=X;
%% processes to eatimate parameters %%
%% initial values %%
for i=1:state
 sigmaJ1(:,i)=Lambda1*X(i)*X2(1)*X2;
 sigmaJ2(:,i)=Lambda1*X(i)*X2(2)*X2;
 sigmaJ3(:,i)=Lambda1*X(i)*X2(3)*X2;
 sigma0(:,i)=Lambda1*X(i)*X2;
 sigmaT1(:,i)=Lambda1*X(i)*interestrate(1)*X2;
 sigmaT2(:,i)=Lambda1*X(i)*interestrate(1)^2*X2;
 sigmaT3(:,i)=Lambda1*X(i)*interestrate(1)*interestrate(2)*X2;
end
%% initial count settings for the batches %%
inta=2;
inte=3+batch;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%% main for-loop: %%
%% parameters are updated 144 times in batches of twenty data points %%
sigmaX(:,1)=X;
SumX(1)=sum(sigmaX(:,1));
for u=1:no_passes
 IR=interestrate(inta:inte);
 n=length(IR);
%% processes for the parameter estimation %%
 for k=1:batch+1
 %Gammai
   Xk=sigmaX(:,inta+k-2);
   for i=1:state
        Gamma(k,i)=exp(-(IR(k+1)*(alpha(i)*IR(k)+gamma(i))/...
                (nu(i)^2))-((alpha(i)*IR(k)+gamma(i))^2/(2*nu(i)^2)));
   end
%% update for markov chain sigmaX(k) %%
   sigmaXk(:,inta+k-1)=Gamma(k,1)*sigmaX(1,inta+k-2)*Pi(:,1)+...
            Gamma(k,2)*sigmaX(2,inta+k-2)*Pi(:,2)+...
            Gamma(k,3)*sigmaX(3,inta+k-2)*Pi(:,3);
   SumX(inta+k-1)=sum(sigmaXk(:,inta+k-1));
   sigmaX(:,inta+k-1)=sigmaXk(:,inta+k-1)/SumX(inta+k-1);
 %% process sigma(J1X) %%
   for r=1:state
        sumJ1(:,r)=Gamma(k,1)*sigmaJ1(1,r)*Pi(:,1)+...
                Gamma(k,2)*sigmaJ1(2,r)*Pi(:,2)+...
                Gamma(k,3)*sigmaJ1(3,r)*Pi(:,3);
        sigmaJ1(:,r)=sumJ1(:,r)+Gamma(k,r)*sigmaX(r,inta+k-2)*Pi(1,r)*E(:,1);
   end
   SumJ1a=sum(sigmaJ1)/SumX(inta+k-2);
   remainderSJ1(:,inta+k-1)=SumJ1a';
 %% process sigma(J2X) %%
   for r=1:state
        sumJ2(:,r)=Gamma(k,1)*sigmaJ2(1,r)*Pi(:,1)+...
                Gamma(k,2)*sigmaJ2(2,r)*Pi(:,2)+...
                Gamma(k,3)*sigmaJ2(3,r)*Pi(:,3);
        sigmaJ2(:,r)=sumJ2(:,r)+Gamma(k,r)*sigmaX(r,inta+k-2)*Pi(2,r)*E(:,2);
   end
   SumJ2a=sum(sigmaJ2)/SumX(inta+k-2);
   remainderSJ2(:,inta+k-1)=SumJ2a';
```

```
%% process sigma(J3X) %%
  for r=1:state
      sumJ3(:,r)=Gamma(k,1)*sigmaJ3(1,r)*Pi(:,1)+...
              Gamma(k,2)*sigmaJ3(2,r)*Pi(:,2)+...
              Gamma(k,3)*sigmaJ3(3,r)*Pi(:,3);
      sigmaJ3(:,r)=sumJ3(:,r)+Gamma(k,r)*sigmaX(r,inta+k-2)*Pi(3,r)*E(:,3);
  end
  SumJ3a=sum(sigmaJ3)/SumX(inta+k-2);
  remainderSJ3(:,inta+k-1)=SumJ3a';
%% process sigma(0) %%
  for r=1:state
      sum0(:,r)=Gamma(k,1)*sigma0(1,r)*Pi(:,1)+...
          Gamma(k,2)*sigma0(2,r)*Pi(:,2)+...
          Gamma(k,3)*sigma0(3,r)*Pi(:,3);
      sigma0(:,r)=sum0(:,r)+Gamma(k,r)*sigmaX(r,inta+k-2)*Pi(:,r);
  end
  Sum0a=sum(sigma0)/SumX(inta+k-2);
  remainderS0(:,inta+k-1)=Sum0a';
%% process sigma(T1) %%
  for r=1:state
      sumT1(:,r)=Gamma(k,1)*sigmaT1(1,r)*Pi(:,1)+...
          Gamma(k,2)*sigmaT1(2,r)*Pi(:,2)+...
          Gamma(k,3)*sigmaT1(3,r)*Pi(:,3);
      sigmaT1(:,r)=sumT1(:,r)+Gamma(k,r)*sigmaX(r,inta+k-2)*IR(k)*Pi(:,r);
  end
  SumT1a=sum(sigmaT1)/SumX(inta+k-2);
  remainderST1(:,inta+k-1)=SumT1a';
%% process sigma(T2) %%
  for r=1:state
      sumT2(:,r)=Gamma(k,1)*sigmaT2(1,r)*Pi(:,1)+...
          Gamma(k,2)*sigmaT2(2,r)*Pi(:,2)+...
          Gamma(k,3)*sigmaT2(3,r)*Pi(:,3);
      sigmaT2(:,r)=sumT2(:,r)+Gamma(k,r)*sigmaX(r,inta+k-2)*IR(k)^2*Pi(:,r);
  end
  SumT2a=sum(sigmaT2)/SumX(inta+k-2);
  remainderST2(:,inta+k-1)=SumT2a';
%% process sigma(T3) %%
  for r=1:state
      sumT3(:,r)=Gamma(k,1)*sigmaT3(1,r)*Pi(:,1)+...
```

```
            Gamma(k,2)*sigmaT3(2,r)*Pi(:,2)+...
            Gamma(k,3)*sigmaT3(3,r)*Pi(:,3);
        sigmaT3(:,r)=sumT3(:,r)+Gamma(k,r)*sigmaX(r,inta+k-2)*...
            interestrate(inta+k-2)*IR(k)*Pi(:,r);
   end
   SumT3a=sum(sigmaT3)/SumX(inta+k-2);
   remainderST3(:,inta+k-1)=SumT3a';
  end
%%%%%%%%%%%%% forecast of interest rate %%%%%%%%%%%%%%%%
    %% parameter for the process %%
 for i=1:state
   beta(i)=gamma(i)/(1-alpha(i));
   a(i)=log(alpha(i))/(u+1-u);
   theta(i)=(nu(i)*sqrt(2*a(i)))/sqrt(1-exp(-2*a(i)*(u+1-u)));
 end
 for k=1:batch
   Xk1=Pi*sigmaX(:,inta+k-2);
   alphaIP1=0;
   gammaIP1=0;
   nuIP1=0;
   for i=1:state
        a11=alpha(i)*Xk1(i);
        alphaIP1=alphaIP1+a11;
        b11=gamma(i)*Xk1(i);
        gammaIP1=gammaIP1+b11;
        c1=nu(i)*Xk1(i);
        nuIP1=nuIP1+c1;
   end
   Xk2=sigmaX(:,inta+k-2);
   alphaIP2=0;
   gammaIP2=0;
   nuIP2=0;
   for i=1:state
        a21=alpha(i)*Xk2(i);
        alphaIP2=alphaIP2+a21;
        b21=gamma(i)*Xk2(i);
        gammaIP2=gammaIP2+b21;
        c21=nu(i)*Xk2(i);
        nuIP2=nuIP2+c21;
```

```
    end
    intr(k+1)=alphaIP1*interestrate(inta+k-2)+gammaIP1;
    LogLikelihood(k)=-0.5*log(2*pi*nuIP1^2)-...
        (interestrate(inta+k-1)-alphaIP1*interestrate(inta+k-2)-gammaIP1)^2/...
        (2*nuIP1^2);
  end
  LogLike(:,u)=LogLikelihood(1:batch)';
  F(:,u)=intr(2:batch+1)';
  x1=sigmaX(1,inta+batch-1);
  x2=sigmaX(2,inta+batch-1);
  x3=sigmaX(3,inta+batch-1);
  AL(u,:)=alpha;
  GA(u,:)=gamma;
  NU(u,:)=nu;
  PROB1(u,:)=Pi(1,1);
  PROB2(u,:)=Pi(1,2);
  PROB3(u,:)=Pi(1,3);
  PROB4(u,:)=Pi(2,1);
  PROB5(u,:)=Pi(2,2);
  PROB6(u,:)=Pi(2,3);
  PROB7(u,:)=Pi(3,1);
  PROB8(u,:)=Pi(3,2);
  PROB9(u,:)=Pi(3,3);
%% optimal estimates for the parameters %%
%% transition probability %%
  for r=1:state
    Pi(1,r)=remainderSJ1(r,inta+batch-1)/remainderS0(r,inta+batch-1);
    Pi(2,r)=remainderSJ2(r,inta+batch-1)/remainderS0(r,inta+batch-1);
    Pi(3,r)=remainderSJ3(r,inta+batch-1)/remainderS0(r,inta+batch-1);
  end
%% alpha %%
  for r=1:state
    alpha(r)=(remainderST3(r,inta+batch-1)-...
        remainderST1(r,inta+batch-1)*gamma(r))/...
        remainderST2(r,inta+batch-1);
  end
%% gamma %%
  for r=1:state
    gamma(r)=(remainderST1(r,inta+batch)-...
```

```
        remainderST1(r,inta+batch-1)*alpha(r))/...
        remainderSO(r,inta+batch-1);
  end
 %% nu %%
  for r=1:state
    nu(r)=sqrt(abs((remainderST2(r,inta+batch)+...
      alpha(r)^2*remainderST2(r,inta+batch-1)+...
      gamma(r)^2*remainderSO(r,inta+batch-1)-...
      2*alpha(r)*remainderST3(r,inta+batch-1)-...
      2*gamma(r)*remainderST1(r,inta+batch)-...
      2*alpha(i)*gamma(i)*remainderST1(r,inta+batch-1))/...
      remainderSO(r,inta+batch-1)));
  end
  inta=inta+batch;
  inte=inte+batch;
 end
 %% Log-likelihood %%
 LogLikeModel=LogLike(:);
 summeLog=sum(LogLike)
 %% Akaike Information Criterion %%
 for i=1:no_passes
     LogLikelihoodTotal(i)=summeLog(i)
     AIC3st(i)=-2*LogLikelihoodTotal(i)+18
 end
 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
 %% plots %%
 PROB=[PROB1 PROB2 PROB3 PROB4 PROB5 PROB6 PROB7 PROB8 PROB9];
 F42=F(:);
 m=length(F42);
 forecast=F42(1:m);
 K=(1:m);
 observation=interestrate(2:m+1);
 subplot(321)
     plot(K, forecast, K, observation)
     ylabel('interest rate')
     xlabel('no of passes')
     title('Daily 1-month T-bill rates 1973-2007')
 subplot(322)
     plot(PROB)
```

```
     ylabel('transition probabilitites')
    xlabel('no of passes')
subplot(323)
    plot(AL)
    ylabel('alpha')
    xlabel('no of passes')
subplot(324)
    plot(GA)
    ylabel('gamma')
    xlabel('no of passes')
subplot(325)
    plot(NU)
     ylabel('nu')
    xlabel('no of passes')
%%%%%%% Error Analysis %%%%%%%
residual=observation-forecast;
%% first difference %%
for f=1:(m-1)
    diffobservation(f)=observation(f+1)-observation(f);
    diffforecast(f)=forecast(f+1)-forecast(f);
end diffresidual=diffobservation-diffforecast;
a=forecast; N=length(a);
b=observation;
c=interestrate(1:m);
%% MSE %%
for i=1:N
diff(i)=(b(i)-a(i))^2;
end
MSE=(1/(N-1))*sum(diff)
%% RAE %%
for i=1:N
absdiff1(i)=abs(a(i)-b(i));
absdiff2(i)=abs(c(i)-b(i));
RAE(i)=absdiff1(i)/absdiff2(i);
end
SRAE=sort(RAE);
s1=N/2;
s2=(N/2)+1;
MdRAE=(SRAE(s1)+SRAE(s2))/2
```

```
cumRAE=sum(absdiff1)/sum(absdiff2);
%% APE %%
for i=1:N diff1(i)=a(i)-b(i);
APE(i)=abs(diff1(i)/b(i)); end SAPE=sort(APE);
s1=N/2; s2=(N/2)+1; MdAPE=(SAPE(s1)+SAPE(s2))/2
-----------------------------------------------
```

Appendix D

Appendix for Chapter 6

D.1 Optimal parameter estimate for μ_Z

We define a new measure $\hat{P}$ by

$$\frac{d\hat{P}}{dP}\Big|_{\mathcal{F}_k} = \Lambda_k^{\mu_Z} = \prod_{l=1}^{k} \lambda_l^{\mu_Z}$$

where

$$\begin{aligned}
\lambda_l^{\mu_Z} &= \frac{\exp\left[-\frac{1}{2\zeta_Z^2}[Z_{l+1} - \hat{\mu}_Z - \mu_B p e^{-\kappa\Delta}]^2\right]}{\exp\left[-\frac{1}{2\zeta_Z^2}[Z_{l+1} - \mu_Z - \mu_B p e^{-\kappa\Delta}]^2\right]} \\
&= \exp\Big[\frac{1}{2\zeta_Z^2}(-(Z_{l+1} - \hat{\mu}_Z - \mu_B p e^{-\kappa\Delta})^2 \\
&\quad +(Z_{l+1} - \mu_Z - \mu_B p e^{-\kappa\Delta})^2)\Big]. \qquad \text{(D.1)}
\end{aligned}$$

The log-likelihood for $\Lambda_k^{\mu_Z}$ is

$$\begin{aligned}
\log \Lambda_k^{\mu_Z} &= \sum_{l=1}^{k}\Big[-\frac{1}{2\zeta_Z^2}(-\hat{\mu}_Z^2 + 2Z_{l+1}\hat{\mu}_Z - 2\hat{\mu}_Z\mu_B p e^{-\kappa\Delta} \\
&\quad +\mu_Z^2 - 2Z_{l+1}\mu_Z + 2\mu_Z\mu_B p e^{-\kappa\Delta})\Big]. \qquad \text{(D.2)}
\end{aligned}$$

We substitute the processes of the Markov chain $\mathbf{x}$ into this log-likelihood and get

$$\begin{aligned}
\log \Lambda_k^{\mu_Z} &= \sum_{i=1}^{n}\Big[-\frac{1}{2\zeta_{Z_i}^2}[\hat{\mu}_{Z_i}^2 O_k^i - 2T_k^i(Z_{k+1})\hat{\mu}_{Z_i} \\
&\quad +2O_k^i\hat{\mu}_{Z_i}\mu_{B_i} p e^{-\kappa_i\Delta} + R(\mu_Z)]\Big] \qquad \text{(D.3)}
\end{aligned}$$

where $R(\mu_Z)$ is a remainder without $\hat{\mu}$. Now, the conditional expectation of the log-likelihood $L(\hat{\mu}_{Z_i}) = E\big[\log \Lambda_k \mid \mathcal{F}_k^Z\big]$is considered. For any process H write $\hat{H}_l = E[H_l \mid \mathcal{F}_k^Z]$.

$$\begin{aligned} L(\hat{\mu}_{Z_i}) &= \sum_{i=1}^{n}\Big[-\frac{1}{2\zeta_{Z_i}^2}\big[\hat{O}_k^i\hat{\mu}_{Z_i}^2 - 2\hat{T}_k^i(Z_{k+1})\hat{\mu}_{Z_i} \\ &\quad +2\hat{O}_k^i\hat{\mu}_{Z_i}\mu_{B_i}pe^{-\kappa_i\Delta} + R(\mu_Z)\big]\Big]. \end{aligned} \tag{D.4}$$

We differentiate $L(\hat{\mu}_{Z_i})$ in $\hat{\mu}_{Z_i}$ and equate the result to 0. This gives

$$\begin{aligned} 2\hat{O}_k^i\hat{\mu}_{Z_i} - 2\hat{T}_k^i(Z_{k+1}) + 2\hat{O}_k^i\mu_{B_i}pe^{-\kappa_i\Delta} &= 0 \\ \text{or } \hat{\mu}_{Z_i} = \frac{\hat{T}_k^i(Z_{k+1}) - \hat{O}_k^i\mu_{B_i}pe^{-\kappa_i\Delta}}{\hat{O}_k^i} &. \end{aligned} \tag{D.5}$$

D.2 Optimal parameter estimate for μ_B

The new measure is defined by the Radon-Nikodým derivative similar to D.1. So,

$$\begin{aligned} \lambda_l^{\mu_B} &= \frac{\exp\big[-\frac{1}{2\zeta_Z^2}(Z_{l+1} - \mu_Z - \hat{\mu}_B pe^{-\kappa\Delta})^2\big]}{\exp\big[-\frac{1}{2\zeta_Z^2}(Z_{l+1} - \mu_Z - \mu_B pe^{-\kappa\Delta})^2\big]} \\ &= \exp\Big[-\frac{1}{2\zeta_Z^2}(\hat{\mu}_B^2p^2e^{-2\kappa\Delta} - 2Z_{l+1}\hat{\mu}_B pe^{-\kappa\Delta} + 2\mu_Z\hat{\mu}_B pe^{-\kappa\Delta} + R(\mu_B))\Big] \end{aligned} \tag{D.6}$$

where $R(\mu_B)$ is a remainder without $\hat{\mu}_B$. We calculate the log-likelihood and include the processes that involves the Markov chain $\mathbf{x}$:

$$\begin{aligned} \log \Lambda_k^{\mu_B} &= \sum_{i=1}^{n}\Big[-\frac{1}{2\zeta_{Z_i}^2}\big[O_k^i\hat{\mu}_{B_i}^2p^2e^{-2\kappa_i\Delta} - 2T_k^i(Z_{k+1})\hat{\mu}_{B_i}pe^{-\kappa_i\Delta} + 2O_k^i\mu_{Z_i}\hat{\mu}_{B_i}pe^{-\kappa_i\Delta} \\ &\quad +R(\mu_B)\big]\Big]. \end{aligned} \tag{D.7}$$

Write $\tilde{H}_l = E[H_l \mid \mathcal{F}_k^Z]$. Now, differentiate the log-likelihood in $\hat{\mu}_{B_i}$ and equate the differential to 0 and get

$$\begin{aligned} &2\tilde{O}_k^i\hat{\mu}_{B_i}p^2e^{-2\kappa_i\Delta} - 2\tilde{T}_k^i(Z_{k+1})pe^{-\kappa_i\Delta} + 2\tilde{O}_k^i\mu_{Z_i}pe^{-\kappa_i\Delta} = 0 \\ \text{or} \quad &\hat{\mu}_{B_i} = \frac{\tilde{T}_k^i(Z_{k+1}) - \tilde{O}_k^i\mu_{Z_i}}{\tilde{O}_k^ipe^{-\kappa_i\Delta}}. \end{aligned} \tag{D.8}$$

D.3 Optimal parameter estimate for β

We define the new measure $\hat{P}$ by

$$\frac{d\hat{P}}{dP}\Big|_{\mathcal{F}_k} = \Lambda_k^\beta = \prod_{l=1}^{k} \lambda_l^\beta \tag{D.9}$$

where

$$\begin{aligned}
\lambda_l^\beta &= \frac{\exp\Big[-\frac{1}{2\zeta_Z^2}\big[Z_{l+1} - \hat{\beta} - [Z_l - \hat{\beta}]e^{-\kappa\Delta} - \mu_B p e^{-\kappa\Delta}\big]^2\Big]}{\exp\Big[-\frac{1}{2\zeta_Z^2}\big[Z_{l+1} - \beta - [Z_l - \beta]e^{-\kappa\Delta} - \mu_B p e^{-\kappa\Delta}\big]^2\Big]} \\
&= \exp\Big[-\frac{1}{2\zeta_Z^2}\Big[\big(Z_{l+1}^2 + \hat{\beta}^2 + ([Z_l - \hat{\beta}]e^{-\kappa\Delta})^2 + (\mu_B p e^{-\kappa\Delta})^2 - 2Z_{l+1}\hat{\beta} \\
&\quad -2Z_{l+1}[Z_l - \hat{\beta}]e^{-\kappa\Delta} - 2Z_{l+1}\mu_B p e^{-\kappa\Delta} - 2\hat{\beta}[Z_l - \hat{\beta}]e^{-\kappa\Delta} \\
&\quad +2\hat{\beta}\mu_B p e^{-\kappa\Delta} + 2[Z_l - \hat{\beta}]e^{-\kappa\Delta}\mu_B p e^{-\kappa\Delta}\big) - \big(Z_{l+1}^2 + \beta^2 + ([Z_l - \beta]e^{-\kappa\Delta})^2 \\
&\quad +(\mu_B p e^{-\kappa\Delta})^2 - 2Z_{l+1}\beta - 2Z_{l+1}[Z_l - \beta]e^{-\kappa\Delta} - 2Z_{l+1}\mu_B p e^{-\kappa\Delta} \\
&\quad -2\beta[Z_l - \beta]e^{-\kappa\Delta} + 2\beta\mu_B p e^{-\kappa\Delta} + 2[Z_l - \beta]e^{-\kappa\Delta}\mu_B p e^{-\kappa\Delta}\big)\Big]\Big].
\end{aligned} \tag{D.10}$$

For the log-likelihood we have

$$\begin{aligned}
\log\Lambda_k^\beta &= \sum_{l=1}^{k}\Big[-\frac{1}{2\zeta_Z^2}\big[\hat{\beta}^2(1 + e^{-2\kappa\Delta} + 2e^{-\kappa\Delta}) + 2\hat{\beta}(-e^{-2\kappa\Delta}Z_l - Z_{l+1} \\
&\quad +Z_{l+1}e^{-\kappa\Delta} - Z_l e^{-\kappa\Delta} + \mu_B p e^{-\kappa\Delta} - e^{-2\kappa\Delta}\mu_B p) + R(\beta)\big]\Big]
\end{aligned} \tag{D.11}$$

where $R(\beta)$ is a remainder which does not include $\hat{\beta}$ terms. We substitute the expressions including the Markov chain with the defined processes $O_k^i = \sum_{l=1}^{k}\langle \mathbf{x}_l, \mathbf{e}_i\rangle$ and $T_k^i(f) = \sum_{l=1}^{k}\langle \mathbf{x}_l, \mathbf{e}_i\rangle f(Z_l)$. Therefore

$$\begin{aligned}
\log\Lambda_k^\beta &= \sum_{i=1}^{n}\Big[-\frac{1}{2\zeta_{Z_i}^2}\Big[\hat{\beta}_i^2\big(O_k^i(1 + e^{-2\kappa_i\Delta} + 2e^{-\kappa_i\Delta}\big) + 2\hat{\beta}_i\big(-T_k^i(Z_l)e^{-2\kappa_i\Delta} \\
&\quad -T_k^i(Z_{l+1}) + T_k^i(Z_{l+1})e^{-\kappa_i\Delta} - T_k^i(Z_l)e^{-\kappa\Delta} \\
&\quad +O_k^i(-e^{-2\kappa_i\Delta}\mu_{Bi}p + \mu_{Bi}pe^{-\kappa_i\Delta})\big)\Big] + R(\beta)\Big].
\end{aligned} \tag{D.12}$$

To calculate the expectation of the log-likelihood conditional on $\mathcal{F}_k^Z$ we set $\tilde{H}_l = E[H_l \mid \mathcal{F}_k^Z]$ for any process H. We differentiate $L(\hat{\boldsymbol{\beta}})$ in $\hat{\beta}_i$ and equate the result to 0. This gives

$$\begin{aligned}
&2\hat{\beta}_i\tilde{O}_k^i(1 + 2e^{-\kappa_i\Delta} + e^{-2\kappa_i\Delta}) - 2(\tilde{T}_k^i(Z_l)(e^{-2\kappa\Delta} + e^{-\kappa_i\Delta}) - 2\tilde{T}_k^i(Z_{l+1})(1 - e^{-\kappa_i\Delta}) \\
&+2\tilde{O}_k^i(-e^{-2\kappa_i\Delta}\mu_{Bi}p + \mu_{Bi}pe^{-\kappa_i\Delta}) = 0\,.
\end{aligned}$$

Henceforth

$$\begin{aligned}\hat{\beta}_i &= \frac{\tilde{T}_k^i(Z_l)(e^{-2\kappa_i\Delta}+e^{-\kappa_i\Delta})+\tilde{T}_k^i(Z_{l+1})(1-e^{-\kappa_i\Delta})}{\tilde{O}_k^i(1+2e^{-\kappa_i\Delta}+e^{-2\kappa_i\Delta l})} \\ &\quad -\frac{\tilde{O}_k^i(-e^{-2\kappa_i\Delta}\mu_{Bi}p+\mu_{Bi}pe^{-\kappa_i\Delta})}{\tilde{O}_k^i(1+2e^{-\kappa_i\Delta}+e^{-2\kappa_i\Delta l})}. \end{aligned} \tag{D.13}$$

D.4 Optimal parameter estimate for ζ_Z

For the MLE of the variance of the Ornstein-Uhlenbeck component in the observation process, we define the Radon-Nikodým derivative $\frac{d\hat{P}}{dP}$ with following λ^{ζ_Z}

$$\begin{aligned}\lambda_l^{\zeta_Z} &= \frac{\frac{1}{\sqrt{\hat{\zeta}_Z^2+\zeta_B^2e^{-2\kappa\Delta}p}}\exp\left[\frac{1}{2(\hat{\zeta}_Z^2+\zeta_B^2e^{-2\kappa\Delta}p)}(Z_{l+1}-\mu_Z-\mu_Bpe^{-\kappa\Delta})^2\right]}{\frac{1}{\sqrt{\zeta_Z^2+\zeta_B^2e^{-2\kappa\Delta}p}}\exp\left[\frac{1}{2(\zeta_Z^2+\zeta_B^2e^{-2\kappa\Delta}p)}(Z_{l+1}-\mu_Z-\mu_Bpe^{-\kappa\Delta})^2\right]} \\ &= \frac{\sqrt{\zeta_Z^2+\zeta_B^2e^{-2\kappa\Delta}p}}{\sqrt{\hat{\zeta}_Z^2+\zeta_B^2e^{-2\kappa\Delta}p}}\exp\Big[-\frac{1}{2(\hat{\zeta}_Z^2+\zeta_B^2e^{-2\kappa\Delta}p)}(Z_{l+1}-\mu_Z-\mu_Bpe^{-\kappa\Delta})^2 \\ &\quad +\frac{1}{2(\zeta_Z^2+\zeta_B^2e^{-2\kappa\Delta}p)}(Z_{l+1}-\mu_Z-\mu_Bpe^{-\kappa\Delta})^2)\Big].\end{aligned}$$

The log-likelihood of $\Lambda_k^{\zeta_Z}$ is therefore

$$\begin{aligned}\log\Lambda_k^{\zeta_Z} &= \sum_{k=1}^{l}\Big(\frac{1}{2}\log\frac{\zeta_Z^2+\zeta_B^2e^{-2\kappa\Delta}p}{\hat{\zeta}_Z^2+\zeta_B^2e^{-2\kappa\Delta}p}-\frac{1}{2(\hat{\zeta}_Z^2+\zeta_B^2e^{-2\kappa\Delta}p)}(Z_{l+1}-\mu_Z-\mu_Bpe^{-\kappa\Delta})^2 \\ &\quad +R(\zeta_Z^2)\Big).\end{aligned} \tag{D.14}$$

Since ζ_Z is regime-switching, we have the following conditional expectation of the log-likelihood including the Markov chain $\mathbf{x}$:

$$\begin{aligned}L(\hat{\zeta}_Z^2) &= \sum_{l=1}^{k}\sum_{i=1}^{n}\Big(-\frac{1}{2}\langle\mathbf{x}_l,\mathbf{e}_i\rangle\log(\hat{\zeta}_{Z_i}^2+\zeta_{B_i}^2e^{-2\kappa_i\Delta}p) \\ &\quad -\langle\mathbf{x}_l,\mathbf{e}_i\rangle\frac{1}{2(\zeta_{Z_i}^2+\zeta_{B_i}^2e^{-2\kappa_i\Delta}p)}(Z_{l+1}-\mu_{Z_i}-\mu_{B_i}pe^{-\kappa_i\Delta})^2\Big)+R(\zeta_Z^2) \\ &= \sum_{i=1}^{n}\Big(-\frac{1}{2}\tilde{O}_k^i\log(\hat{\zeta}_{Z_i}^2+\zeta_{B_i}^2e^{-2\kappa_i\Delta}p) \\ &\quad -\frac{1}{2(\hat{\zeta}_{Z_i}^2+\zeta_{B_i}^2e^{-2\kappa_i\Delta}p)}(\tilde{T}_k^i(Z_{k+1}^2)+\tilde{O}_k^i\mu_{Z_i}^2+\tilde{O}_k^i\mu_{B_i}^2p^2e^{-2\kappa_i\Delta}-2\tilde{T}_k^i(Z_{k+1})\mu_{Z_i} \\ &\quad -2\tilde{T}_k^i(Z_{k+1})\mu_{B_i}pe^{-\kappa_i\Delta}+2\tilde{O}_k^i\mu_{Z_i}\mu_{B_i}pe^{-\kappa_i\Delta})\Big)+R(\zeta_Z^2).\end{aligned} \tag{D.15}$$

To find the maximum we differentiate $L(\hat{\zeta}_Z^2)$ with respect to each $\hat{\zeta}_{Z_i}^2$ and equate the resulting derivative to 0. We then have

$$\tilde{T}_k^i(Z_{k+1}^2) + \tilde{O}_k^i(\mu_{Z_i}^2 + \mu_{B_i}^2 p^2 e^{-2\kappa_i\Delta} + 2\mu_{Z_i}\mu_{B_i} p e^{-\kappa_i\Delta}) - 2\tilde{T}_k^i(Z_{k+1})(\mu_{Z_i} + \mu_{B_i} p e^{-\kappa_i\Delta})$$
$$= \tilde{O}_k^i(\hat{\zeta}_{Z_i}^2 + \zeta_{B_i}^2 e^{-2\kappa_i\Delta} p)\,. \tag{D.16}$$

Consequently, we find that the optimal parameter estimate for $\hat{\zeta}_{Z_i}^2$ is

$$\hat{\zeta}_{Z_i}^2 = \frac{\tilde{T}_k^i(Z_{k+1}^2) + \tilde{O}_k^i(\mu_{Z_i}^2 + \mu_{B_i}^2 p^2 e^{-2\kappa_i\Delta} + 2\mu_{Z_i}\mu_{B_i} p e^{-\kappa_i\Delta} - \zeta_{B_i}^2 p e^{-2\kappa_i\Delta})}{\tilde{O}_k^i}$$
$$- \frac{2\tilde{T}_k^i(Z_{k+1})(\mu_{Z_i} + \mu_{B_i} e^{-\kappa_i\Delta} p)}{\tilde{O}_k^i}\,. \tag{D.17}$$

D.5 Optimal parameter estimate for ζ_B^2

The Radon-Nikodým derivative $\frac{d\hat{P}}{dP}$ is defined by

$$\lambda_l^{\zeta_B} = \sqrt{\frac{\zeta_Z^2 + \zeta_B^2 p e^{-2\kappa\Delta}}{\zeta_Z^2 + \hat{\zeta}_B^2 e^{-2\kappa\Delta} p}} \exp\Big[-\frac{1}{2(\zeta_Z^2 + \hat{\zeta}_B^2 e^{-2\kappa\Delta} p)}(Z_{l+1} - \mu_Z - \mu_B p e^{-\kappa\Delta})^2$$
$$+ \frac{1}{2(\zeta_Z^2 + \zeta_B^2 e^{-2\kappa\Delta} p)}(Z_{l+1} - \mu_Z - \mu_B p e^{-\kappa\Delta})^2\Big]. \tag{D.18}$$

Therefore the log-likelihood of $\Lambda_k^{\zeta_B}$ is given by

$$\log \Lambda_k^{\zeta_B} = \sum_{l=1}^{k}\Big(-\frac{1}{2}\log(\zeta_Z^2 + \hat{\zeta}_B^2 e^{-2\kappa\Delta} p)$$
$$- \frac{1}{2(\zeta_Z^2 + \hat{\zeta}_B^2 e^{-2\kappa\Delta} p)}(Z_{l+1} - \mu_Z - \mu_B p e^{-\kappa\Delta})^2 + R(\zeta_B^2 p)\Big). \tag{D.19}$$

From (D.19), we obtain the conditional expectation

$$L(\hat{\zeta}_B^2) = \sum_{i=1}^{n}\Big(-\frac{1}{2}\tilde{O}_k^i \log(\zeta_{z_i}^2 + \hat{\zeta}_{B_i}^2 e^{-2\kappa_i\Delta} p) - \frac{1}{2(\zeta_{Z_i}^2 + \hat{\zeta}_{B_i}^2 e^{-2\kappa_i\Delta} p)}\big(\tilde{T}_k^i(Z_{k+1}^2)$$
$$+ \tilde{O}_k^i(\mu_{Z_i}^2 + \mu_{B_i}^2 p^2 e^{-2\kappa_i\Delta} + 2\mu_{Z_i}\mu_{B_i} p e^{-\kappa_i\Delta})$$
$$- 2\tilde{T}_k^i(Z_{k+1})(\mu_{Z_i} + \mu_{B_i} p e^{-\kappa_i\Delta})\big)\Big) + R(\zeta_B^2 p). \tag{D.20}$$

Differentiating $(D.20)$ with respect to $\hat{\zeta}^2_{B_i}$ gives

$$\hat{\zeta}^2_{B_i} = \frac{\tilde{T}^i_k(Z^2_{k+1}) + \tilde{O}^i_k(\mu^2_{Z_i} + \mu^2_{B_i}p^2e^{-2\kappa_i\Delta} + 2\mu_{Z_i}\mu_{B_i}pe^{-\kappa_i\Delta} - \zeta^2_{Z_i})}{\tilde{O}^i_k pe^{-2\kappa_i\Delta}} - \frac{2\tilde{T}^i_k(Z_{k+1})(\mu_{Z_i} + \mu_{B_i}pe^{-\kappa_i\Delta})}{\tilde{O}^i_k pe^{-2\kappa_i\Delta}} \quad (D.21)$$

D.6 MATLAB source code: Implementation of the 3-state HMM for electricity spot prices

The model for forecasting daily electricity spot prices is implemented in MATLAB. Here, the source code for the implementation of the 3-state hidden Markov model is stated. The main body includes the filtering and parameter estimation together with the one-step ahead forecast. The error analysis and the calculation of the expected spot at delivery is stated in the last part of this code. The implementation for other number of states is done accordingly.

```
%% Hidden Markov model for daily electricity spot prices %%
%% data: daily average prices in NOK/MwH  (1360 data points (1998-2002)) %%
%% data is log of deseasonalised spotprice %%
clear;
%% deseasonlised data %%
load logremaind
%% simulated muB(i)+ zetaB(i)* randn %%
load remainderJump3st
%% seasonal function D(k) %%
load det
%% read in original data for electricity price %%
price=logremaind;
%% states of markov chain %%
```

```
state = 3;
%% no of algorithm passes %%
no_update=34
%% batch interval %%
batch=40;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% initial values for parameters %%
beta=[-0.1 -0.15 -0.2]
zetaZ=[1.5*std(price) std(price) 2*std(price)];
muZ=[mean(price) 2*mean(price) 4*mean(price)];
muB=[-0.5 0 0.5]
zetaB=[0.8 1 0.1];
kappa=[300 100 200];
zeta=[sqrt(abs((2*kappa(1)*zetaZ(1)^2)/(1-exp(-2*kappa(1)*(1/365))))),...
        sqrt(abs((2*kappa(2)*zetaZ(2)^2)/(1-exp(-2*kappa(2)*(1/365))))),...
        sqrt(abs((2*kappa(3)*zetaZ(3)^2)/(1-exp(-2*kappa(3)*(1/365)))))];
Pi=[1/3 1/3 1/3; 1/3 1/3 1/3; 1/3 1/3 1/3];
%% initial value for markov chain %%
Z=[0.99 0.005 0.005]';
%% initial values: Markov chain in time 2 equals Markov chain Z in time 1 %%
Z2=Z;
E=eye(state);
%% jump counter k %%
jck=0;
%% initial Radon-Nikodym derivative for change of measure %%
deltat=1/365;
Lambda1=10;
%% initial values for processes of Markov chain Z to estimate parameters %%
for i=1:state
  etaJ1(:,i)=Lambda1*Z(i)*Z2(1)*Z2;
  etaJ2(:,i)=Lambda1*Z(i)*Z2(2)*Z2;
  etaJ3(:,i)=Lambda1*Z(i)*Z2(3)*Z2
end
SumJ1a=sum(etaJ1);
remainderSJ1(:,1)=SumJ1a';
SumJ2a=sum(etaJ2);
remainderSJ2(:,1)=SumJ2a';
SumJ3a=sum(etaJ3);
remainderSJ3(:,1)=SumJ3a';
```

```
for i=1:state
  etaO(:,i)=Lambda1*Z(i)*Z2;
end
SumOa=sum(etaO);
remainderSO(:,1)=SumOa';
for i=1:state
  etaT1(:,i)=Lambda1*Z(i)*price(1)*Z2;
end
SumT1a=sum(etaT1);
remainderST1(:,1)=SumT1a';
for i=1:state
  etaT2(:,i)=Lambda1*Z(i)*price(1)^2*Z2;
end
SumT2a=sum(etaT2);
remainderST2(:,1)=SumT2a';
for i=1:state
  etaT3(:,i)=Lambda1*Z(i)*price(2)*Z2;
end
SumT3a=sum(etaT3);
remainderST3(:,1)=SumT3a';
for i=1:state
  etaT4(:,i)=Lambda1*Z(i)*price(2)*price(2)*Z2;
end
SumT4a=sum(etaT4);
remainderST4(:,1)=SumT4a';
etaZ(:,1)=Z;
etaZk(:,1)=Z;
SumZ(1)=sum(etaZ(:,1));
%% initial count settings for the batches %%
inta=2;
inte=3+batch;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% estimation of constant intensity lambda %%
s=std(price);
leng=length(price);
upperbound=mean(price)+2*s;
lowerbound=mean(price)-2*s;
%% number of jumps in the actual data %%
jck1=0;
```

```
jck2=0;
tau1=zeros(1,leng);
for k=1:leng
  if price(k)<lowerbound
    jck1=jck1+1;
    tau1(k)=1/365;
  end
  if price(k)>upperbound
    jck2=jck2+1;
    tau1(k)=1/365;
  end
end
tau=nonzeros(tau1);
%% jump intensity lambda for a day %%
Nt1=(jck1+jck2)/(no_update*batch)
%% jump intensity lambda for a period of length batch %%
counterbatch=365/batch;
Nt=floor(Nt1*batch);
%% jump intensity for a year %%
Nt2=Nt1*365;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% main for-loop: %%
for u=1:no_update
  taubatch=tau1((u*batch+1-batch):(u*batch));
  %% batches of "batch" data points %%
  spot=price(inta:inte);
  %% how many data points %%
  n=length(spot);
  for zaehler=1:batch+1
    for i=1:state
      %% CPP:compound Poisson process %%
      CPP(i)=0;
      taubatch1=nonzeros(taubatch);
      J(i)=remainderJump3st(inta+zaehler-1,i);
      for c=1:Nt
        CPP(i)=CPP(i)+exp(-kappa(i)*(deltat-taubatch(c)))*J(i);
      end
    end
    %% Gamma_i %%
```

```
Zk=etaZ(:,inta+zaehler-2);
for i=1:state
  d(i)=zeta(i)*sqrt(((1-exp(-2*kappa(i)*deltat))/(2*kappa(i)*deltat)));
  Gamma(zaehler,i)=exp(-(1/d(i)^2)*(spot(zaehler)*exp(-kappa(i)*deltat)...
      +beta(i)*(1-exp(-kappa(i)*deltat))+CPP(i))*spot(zaehler+1)...
      -(1/(2*d(i)^2))*(spot(zaehler)*exp(-kappa(i)*deltat)...
      +beta(i)*(1-exp(-kappa(i)*deltat))+CPP(i))^2);
end
%% update for markov chain etaZ(k) (estimator for the state) %%
etaZk(:,inta+zaehler-1)=Gamma(zaehler,1)*etaZ(1,inta+zaehler-2)*Pi(:,1)...
    +Gamma(zaehler,2)*etaZ(2,inta+zaehler-2)*Pi(:,2)...
    +Gamma(zaehler,3)*etaZ(3,inta+zaehler-2)*Pi(:,3);
SumZ(inta+zaehler-1)=sum(etaZk(:,inta+zaehler-1));
etaZ(:,inta+zaehler-1)=etaZk(:,inta+zaehler-1)/SumZ(inta+zaehler-1);
%% process eta(J1Z) %%
for r=1:state
  sumJ1(:,r)=Gamma(zaehler,1)*etaJ1(1,r)*Pi(:,1)...
      +Gamma(zaehler,2)*etaJ1(2,r)*Pi(:,2)...
      +Gamma(zaehler,3)*etaJ1(3,r)*Pi(:,3);
  etaJ1(:,r)=sumJ1(:,r)+Gamma(zaehler,r)*etaZ(r,inta+zaehler-2)...
      *Pi(1,r)*E(:,1);
end
SumJ1a=sum(etaJ1)/SumZ(inta+zaehler-1);
remainderSJ1(:,inta+zaehler-1)=SumJ1a';
%% process eta(J2Z) %%
for r=1:state
  sumJ2(:,r)=Gamma(zaehler,1)*etaJ2(1,r)*Pi(:,1)...
      +Gamma(zaehler,2)*etaJ2(2,r)*Pi(:,2)...
      +Gamma(zaehler,3)*etaJ2(3,r)*Pi(:,3);
  etaJ2(:,r)=sumJ2(:,r)+Gamma(zaehler,r)*etaZ(r,inta-2+zaehler)...
      *Pi(2,r)*E(:,2);
end
SumJ2a=sum(etaJ2)/SumZ(inta-1+zaehler);
remainderSJ2(:,inta+zaehler-1)=SumJ2a';
%% process eta(J3Z) %%
for r=1:state
  sumJ3(:,r)=Gamma(zaehler,1)*etaJ3(1,r)*Pi(:,1)...
      +Gamma(zaehler,2)*etaJ3(2,r)*Pi(:,2)...
      +Gamma(zaehler,3)*etaJ3(3,r)*Pi(:,3);
```

```
   etaJ3(:,r)=sumJ3(:,r)+Gamma(zaehler,r)*etaZ(r,inta-2+zaehler)...
       *Pi(3,r)*E(:,3);
 end
 SumJ3a=sum(etaJ3)/SumZ(inta-1+zaehler);
 remainderSJ3(:,inta+zaehler-1)=SumJ3a';
 %% process eta(0) %%
 for r=1:state
   sum0(:,r)=Gamma(zaehler,1)*eta0(1,r)*Pi(:,1)...
       +Gamma(zaehler,2)*eta0(2,r)*Pi(:,2)...
       +Gamma(zaehler,3)*eta0(3,r)*Pi(:,3);
   eta0(:,r)=sum0(:,r)+Gamma(zaehler,r)*etaZ(r,inta+zaehler-2)*Pi(:,r);
 end
 Sum0a=sum(eta0)/SumZ(inta+zaehler-1);
 remainderS0(:,inta+zaehler-1)=Sum0a';
 %% process eta(T1) %%
 for r=1:state
   sumT1(:,r)=Gamma(zaehler,1)*etaT1(1,r)*Pi(:,1)...
       +Gamma(zaehler,2)*etaT1(2,r)*Pi(:,2)...
       +Gamma(zaehler,3)*etaT1(3,r)*Pi(:,3);
   etaT1(:,r)=sumT1(:,r)+Gamma(zaehler,r)*etaZ(r,inta+zaehler-2)...
       *spot(zaehler)*Pi(:,r);
 end
 SumT1a=sum(etaT1)/SumZ(zaehler+inta-1);
 remainderST1(:,inta+zaehler-1)=SumT1a';
 %% process eta(T2) %%
 for r=1:state
   sumT2(:,r)=Gamma(zaehler,1)*etaT2(1,r)*Pi(:,1)...
       +Gamma(zaehler,2)*etaT2(2,r)*Pi(:,2)...
       +Gamma(zaehler,3)*etaT2(3,r)*Pi(:,3);
   etaT2(:,r)=sumT2(:,r)+Gamma(zaehler,r)*etaZ(r,inta-2+zaehler)...
       *spot(zaehler)^2*Pi(:,r);
 end
 SumT2a=sum(etaT2)/SumZ(zaehler+inta-1);
 remainderST2(:,inta+zaehler-1)=SumT2a';
 %% process eta(T3) %%
 for r=1:state
   sumT3(:,r)=Gamma(zaehler,1)*etaT3(1,r)*Pi(:,1)...
       +Gamma(zaehler,2)*etaT3(2,r)*Pi(:,2)...
       +Gamma(zaehler,3)*etaT3(3,r)*Pi(:,3);
```

```
    etaT3(:,r)=sumT3(:,r)+Gamma(zaehler,r)*etaZ(r,inta-2+zaehler)...
        *spot(zaehler+1)*Pi(:,r);
  end
  SumT3a=sum(etaT3)/SumZ(inta+zaehler-1);
  remainderST3(:,inta+zaehler-1)=SumT3a';
  %% process eta(T4) %%
  for r=1:state
    sumT4(:,r)=Gamma(zaehler,1)*etaT4(1,r)*Pi(:,1)...
        +Gamma(zaehler,2)*etaT4(2,r)*Pi(:,2)...
        +Gamma(zaehler,3)*etaT4(3,r)*Pi(:,3);
    etaT4(:,r)=sumT4(:,r)+Gamma(zaehler,r)*etaZ(r,inta-2+zaehler)...
        *spot(zaehler+1)^2*Pi(:,r);
  end
  SumT4a=sum(etaT4)/SumZ(inta+zaehler-1);
  remainderST4(:,inta+zaehler-1)=SumT4a';
%% forecast of electricity prices %%
%% parameter for the process %%
Zk=etaZ(:,zaehler+inta-2);
Zk1=Pi*etaZ(:,zaehler+inta-2);
%% scalar products %%
kappaIP=0;
betaIP=0;
zetaIP=0;
muZIP=0;
zetaZIP=0;
muBIP=0;
zetaBIP=0;
for i=1:state
  a1=kappa(i)*Zk1(i);
  kappaIP=kappaIP+a1;
  b1=beta(i)*Zk1(i);
  betaIP=betaIP+b1;
  h1=zeta(i)*Zk1(i);
  zetaIP=zetaIP+h1;
  c1=muZ(i)*Zk1(i);
  muZIP=muZIP+c1;
  d1=zetaZ(i)*Zk1(i);
  zetaZIP=zetaZIP+d1;
  e1=muB(i)*Zk1(i);
```

```
      muBIP=muBIP+e1;
      f1=zetaB(i)*Zk1(i);
      zetaBIP=zetaBIP+f1;
    end
    %% compound poisson process %%
    CPP1=Nt1*exp(-kappaIP*(deltat))*muBIP;
    %% one-step ahead forecast of log return %%
    intr(zaehler+1)=price(inta+zaehler-2)*exp(-kappaIP*deltat)...
        +betaIP*(1-exp(-kappaIP*deltat))+(zaehler)*CPP1;
 end
 F(:,u)=intr(2:batch+1)';
 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
 %% parameter estimates %%
 AL(u,:)=kappa;
 B(u,:)=beta;
 ZETA(u,:)=zeta;
 MUZ(u,:)=muZ;
 MUB(u,:)=muB;
 ZETAZ(u,:)=zetaZ;
 ZETAB(u,:)=zetaB;
 PROB1(u,:)=Pi(1,1);
 PROB2(u,:)=Pi(1,2);
 PROB3(u,:)=Pi(1,3);
 PROB4(u,:)=Pi(2,1);
 PROB5(u,:)=Pi(2,2);
 PROB6(u,:)=Pi(2,3);
 PROB7(u,:)=Pi(3,1);
 PROB8(u,:)=Pi(3,2);
 PROB9(u,:)=Pi(3,3);
 jck=Nt2;
 %% optimal estimations for the parameters %%
 %% transition probability %%
 for r=1:state
   Pi(1,r)=remainderSJ1(r,u)/remainderSO(r,u);
   Pi(2,r)=remainderSJ2(r,u)/remainderSO(r,u);
   Pi(3,r)=remainderSJ3(r,u)/remainderSO(r,u);
 end
 %% beta %%
 for r=1:state
```

```
   beta(r)=(remainderST1(r,u)*(exp(-2*kappa(r)*deltat)...
       +exp(-kappa(r)*deltat))+remainderST3(r,u)*(1-exp(-kappa(r)*deltat))...
       -remainderSO(r,u)*(-exp(-2*kappa(r)*deltat)*muB(r)*jck...
       +exp(-kappa(r)*deltat)*muB(r)*jck))/(remainderSO(r,u)...
       *(1+exp(-2*kappa(r)*deltat)+2*exp(-kappa(r)*deltat)));
end
%% muB %%
for r=1:state
   muB(r)=(remainderST3(r,u)-remainderSO(r,u)*muZ(r))...
       /(remainderSO(r,u)*jck*exp(-kappa(r)*deltat));
end
%% muZ %%
for r=1:state
   muZ(r)=(remainderST3(r,u)-remainderSO(r,u)*muB(r)...
       *exp(-kappa(r)*deltat)*jck)/remainderSO(r,u);
end
%% zetaB %%
for r=1:state
   zetaB(r)=sqrt(abs((remainderST4(r,u)+remainderSO(r,u)...
       *(muZ(r)^2+muB(r)^2*jck^2*exp(-2*kappa(r)*deltat)...
       +2*muZ(r)*muB(r)*exp(-kappa(r)*deltat)*jck-2*zetaZ(r)^2)...
       -2*remainderST3(r,u)*(muZ(r)+muB(r)*exp(-kappa(r)*deltat)*jck))...
       /(2*remainderSO(r,u)*exp(-2*kappa(r)*deltat)*jck)));
end
%% zetaZ %%
for r=1:state
   zetaZ(r)= sqrt(abs((remainderST4(r,u)+remainderSO(r,u)...
       *(muZ(r)^2+muB(r)^2*jck^2*exp(-2*kappa(r)*deltat)...
       +2*muZ(r)*muB(r)*exp(-kappa(r)*deltat)*jck-2*zetaB(r)^2*jck...
       *exp(-2*kappa(r)*deltat))-2*remainderST3(r,u)...
       *(muZ(r)+muB(r)*jck*exp(-kappa(r)*deltat)))/(2*remainderSO(r,u))));
end
%% kappa %%
for r=1:state
   if (muZ(r) > spot(n-1)) & ( spot(n-1)-beta(r) > 0 )
      kappa(r)=-log((muZ(r)-beta(r))/(spot(n-1)-beta(r)))*(1/deltat);
   elseif (muZ(r) < spot(n-1)) & ( spot(n-1)-beta(r) < 0 )
      kappa(r)=-log((muZ(r)-beta(r))/(spot(n-1)-beta(r)))*(1/deltat);
   else
```

```
      kappa(r)=kappa(r);
    end
  end
  %% zeta %%
  for r=1:state
    zeta(r)=sqrt(((2*kappa(r)*deltat*zetaZ(r)^2)...
        /(1-exp(-2*kappa(r)*deltat))));
  end
  inta=inta+batch;
  inte=inte+batch;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Data preparation for plots %%
PROB=[PROB1 PROB2 PROB3 PROB4 PROB5 PROB6 PROB7 PROB8 PROB9];
F42=F(:);
m=length(F42);
forecast=F42(1:m);
K=(1:m);
observation=price(2:m+1);
subplot(331)
  plot(K, forecast, K, observation)
  ylabel('spot price')
  xlabel('time')
  title('Daily electricity spot prices 12/98-07/02')
subplot(332)
  plot(PROB)
  ylabel('probability')
  xlabel('no of passes')
subplot(333)
  plot(B)
  ylabel('beta')
  xlabel('no of passes')
subplot(336)
  plot(AL)
  ylabel('kappa')
  xlabel('no of passes')
subplot(337)
  plot(MUB)
  ylabel('mu B')
```

```
   xlabel('time')
 subplot(338)
   plot(ZETAB)
   ylabel('zeta B')
   xlabel('time')
 subplot(334)
   plot(MUZ)
   ylabel('mu Z')
   xlabel('no of passes')
 subplot(335)
   plot(ZETAZ)
   ylabel('zeta Z')
   xlabel('time')
 subplot(339)
    plot(ZETA)
    ylabel('zeta')
    xlabel('time')
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Error Analysis for deseasonalised log spot prices %%
residual=observation-forecast;
a=forecast;
N=length(a)
b=observation;
%% MSE %%
for x=1:N
  diff(x)=(b(x)-a(x))^2;
end
MSE=(1/(N-1))*sum(diff)
%% MdAPE %%
for i=1:N
  diff1(i)=a(i)-b(i);
  APE(i)=abs(diff1(i)/b(i));
end SAPE=sort(APE); s1=N/2; s2=(N/2)+1;
MdAPE=(SAPE(s1)+SAPE(s2))/2
c1=price(1:m);
%% MdRAE %%
for i=1:N
  absdiff1(i)=abs(a(i)-b(i));
  absdiff2(i)=abs(c1(i)-b(i));
```

```
  RAE(i)=absdiff1(i)/absdiff2(i);
end
SRAE=sort(RAE);
s1=N/2;
s2=(N/2)+1;
MdRAE=(SRAE(s1)+SRAE(s2))/2
%% Error analysis for spot price forecast %%
Forecastspot=exp(forecast).*det(2:1361)';
Observationspot=exp(observation).*det(2:1361)';
%% MSE %%
for x=1:N
  diff(x)=(Forecastspot(x)-Observationspot(x))^2;
end
MSEspot=(1/(N-1))*sum(diff)
%% MdAPE %%
for i=1:N
  diff1(i)=Observationspot(i)-Forecastspot(i);
  APE(i)=abs(diff1(i)/Forecastspot(i));
end SAPE=sort(APE); s1=N/2; s2=(N/2)+1;
MdAPEspot=(SAPE(s1)+SAPE(s2))/2;
c=exp(price(1:m)).*det(1:1360)';
%% MdRAE %%
for i=1:N
  absdiff1(i)=abs(Observationspot(i)-Forecastspot(i));
  absdiff2(i)=abs(c(i)-Forecastspot(i));
  RAE(i)=absdiff1(i)/absdiff2(i);
end
SRAE=sort(RAE);
s1=N/2;
s2=(N/2)+1;
MdRAEspot=(SRAE(s1)+SRAE(s2))/2
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Pricing electricity contracts %%
%% calculating the expected spot price %%
%% with data point 1330 as a starting point %%
for start=1:15
  deterministic=det((1309):1361)';
  SPOT=Observationspot((1308):1360);
  lambda=Nt1;
```

```
  for mat=(start+1):(30+start)
    s=start;
    firstterm=exp(muBIP*exp(-kappaIP*deltat^2*(mat-start)),...
        +0.5*zetaBIP^2*exp(-2*kappaIP*deltat^2*(mat-start)));
    for N=1:((mat-start)*4-1)
        integral(N)=2*exp(exp(-kappaIP*deltat^2*(mat-s))*muBIP,...
            +0.5*zetaBIP^2*exp(-2*kappaIP*deltat^2*(mat-start)));
        s=s+0.25;
    end
    lastterm=exp(exp(-kappaIP*deltat^2*(mat-s))*muBIP,...
        +0.5*zetaBIP^2*exp(-2*kappaIP*deltat^2*(mat-s)));
    laenge=length(integral);
    summeintegral=(mat-start)/(2*(laenge+1))*lambda*(sum(integral),...
        +firstterm+lastterm);
    Forward3st(start,(mat-start))=deterministic(mat),...
        *((SPOT(start)/deterministic(start)),...
        ^exp(-kappaIP*deltat^2*(mat-start))) ,...
        *exp(betaIP*(1-exp(-kappaIP*deltat^2*(mat-start))),...
        +0.5*zetaIP^2*(1-exp(-2*kappaIP*deltat^2*(mat-start)))) ,...
        *exp(summeintegral-lambda*(mat-start));
  end
end
-----------------------------------------
```

Appendix E

Appendix for Chapter 7

E.1 MATLAB source code: Implementation of the asset allocation models

The source code for the asset allocation problem is presented here. Subsection E.1.1 contains the code for the implementation of the 3-state hidden Markov model for a two-dimensional observation process. The main body includes the filtering and parameter estimation together with the generation of one-step ahead forecasts for two time series data, namely the indices NASDAQ and Dow Jones. The MATLAB source code for the calculation of the investment strategies is given in subsection E.1.2.

E.1.1 Matlab code for the 3-state HMM: Case of vector observations

```
-----------------------------------------

%% HMM for two-dimensional observation process %%
%% actual data are indeces %%
clear;
%% workspace with NASDAQ and DOWJONES data %%
%% weekly closing data for ten years (1997 - 2007) %%
```

```
load Nasdaqdaily9707
load Dowjonesdaily9707
%% h: how many steps of forecast %%
for h=1:1
%% actual data %%
C1=NASDAQ9707;
C2=DOW_JONES9707;
%% log-returns (y(k+1)=log(C(k+1)/C(k)) %%
Nasdaqret=price2ret(NASDAQ9707);
Dowjonesret=price2ret(DOW_JONES9707);
%% state of the Markov chain %%
state=3;
%% dimension of vector observation %%
dim=2;
%%%%%%%%%%%%% initial values %%%%%%%%%%%%%%%%%%%%%%%
%% intial values for parameters of observation process %%
%% mu and sigma %%
mean1=mean(Nasdaqret)
std1=std(Nasdaqret)
mean2=mean(Dowjonesret)
std2=std(Dowjonesret)
%% initial values for NASDAQ parameters %%
mu1=[0.5*mean1; mean1; 1.5*mean1];
sigma1=[0.8*std1; std1; 2*std1];
%% initial values for DOW_JONES parameters %%
mu2=[0.5*mean2; mean2; 1.5*mean2];
sigma2=[0.8*std2; std2; 2*std2];
%% initial values for transition probability matrix A %%
A=[1/3 1/3 1/3; 1/3 1/3 1/3; 1/3 1/3 1/3];
%% intial values for conditional expectation of %%
%% Markov chain %%
xi(:,1)=[0.99 0.005 0.005]';
E=eye(state);
%% first p=E[X_1|Y_1] %%
x1=xi(:,1);
nenner1=sum(x1);
p1=x1./nenner1;
xhat(:,1)=p1;
%% Lambda for calculating start values for the recursion of %%
```

```
%% the gammas %%
w=(Nasdaqret(1)-(mu1(1)*x1(1)+mu1(2)*x1(2)+mu1(3)*x1(3)))...
    /(sigma1(1)*x1(1)+sigma1(2)*x1(2)+sigma1(3)*x1(3));
w1=((2*pi)^(-1/2))*exp(-(w^2)/2);
w2=((2*pi)^(-1/2))*exp(-(Nasdaqret(1)^2)/2);
Lambda11=w1/((sigma1(1)*x1(1)+sigma1(2)*x1(2)...
    +sigma1(3)*x1(3))*w2);
v=(Dowjonesret(1)-(mu1(1)*x1(1)+mu1(2)*x1(2)+mu1(3)*x1(3)))...
    /(sigma1(1)*x1(1)+sigma1(2)*x1(2)+sigma1(3)*x1(3));
v1=((2*pi)^(-1/2))*exp(-(v^2)/2);
v2=((2*pi)^(-1/2))*exp(-(Dowjonesret(1)^2)/2);
Lambda21=v1/((sigma1(1)*x1(1)+sigma1(2)*x1(2)...
    +sigma1(3)*x1(3))*v2);
Lambda1=Lambda11*Lambda21
%% first diagonal matrix for first gammas %%
for i=1:state
  Y11(i)=(Nasdaqret(2)-mu1(i)*x1(i))/(sigma1(i)*x1(i));
  %% standard normal distribution of Y %%
  Z11(i)=((2*pi)^(-1/2))*exp(-(Y11(i)^2)/2);
  %% standard normal distribution of Nasdaqret %%
  N11(i)=(sigma1(i)*x1(i))*(((2*pi)^(-1/2))...
      *exp(-(Nasdaqret(2)^2)/2));
  Y21(i)=(Dowjonesret(2)-mu2(i)*x1(i))/(sigma2(i)*x1(i));
  %% standard normal distribution of Y %%
  Z21(i)=((2*pi)^(-1/2))*exp(-(Y21(i)^2)/2);
  %% standard normal distribution of Dowjonesret %%
  N21(i)=(sigma2(i)*x1(i))*(((2*pi)^(-1/2))...
      *exp(-(Dowjonesret(2)^2)/2));
  %% entries for diagonal matrix %%
  D1(i)=(Z11(i)*Z21(i))/(N11(i)*N21(i)) ;
end
%% diagonal matrix with entries D1 on the diagonal %%
DD1=diag(D1);
xi(:,2)=A*DD1*xi(:,1);
x2=xi(:,2);
nenner2=sum(x2);
p2=x2./nenner2;
xhat(:,2)=p2;
%% first gammas for the recursion formulas %%
```

```
for i=1:state
  gammaj1(:,i)=Lambda1*p1(i)*p2(1)*p2;
  gammaj2(:,i)=Lambda1*p1(i)*p2(2)*p2;
  gammaj3(:,i)=Lambda1*p1(i)*p2(3)*p2;
  gammao(:,i)=Lambda1*p1(i)*p2;
  gammatnasdaq(:,i)=Lambda1*p1(i)*Nasdaqret(1)*p2;
  gammatdowjones(:,i)=Lambda1*p1(i)*Dowjonesret(1)*p2;
  gammatqnasdaq(:,i)=Lambda1*p1(i)*Nasdaqret(1)^2*p2;
  gammatqdowjones(:,i)=Lambda1*p1(i)*Dowjonesret(1)^2*p2;
end
%% data points within one batch %%
batch=10;
%% initial values for intervals %%
a=2;
e=batch+a;
%% number of parameter updates %%
no_paramupdate=(floor((length(Nasdaqret)-2)/batch))
%%%%%% main for-loop %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for u=1:no_paramupdate
  %% batches of ten data points of Nasdaq returns %%
  B1=Nasdaqret(a:e);
  %% batches of ten data points from Dow Jones returns %%
  B2=Dowjonesret(a:e);
  n=length(B1);
  %% Calculations for xi %%
  for k=1:batch
    hh=xi(:,a+k-1);
    hh1=sum(hh);
    %% diagonal matrix: entries for fist and second %%
    %% dimension are calculated and then %%
    %% multiplied for entries of diagonal matrix %%
    for i=1:state
      %% entries for first dimension (NASDAQ) %%
      Y1(k,i)=(B1(k+1)-mu1(i)*hh(i))/(sigma1(i)*hh(i));
      Z1(k,i)=((2*pi)^(-1/2))*exp(-(Y1(k,i)^2)/2);
      N1(k,i)=(sigma1(i)*hh(i))*(((2*pi)^(-1/2))...
          *exp(-(B1(k+1)^2)/2));
      %% entries for second dimension (DOWJONES) %%
      Y2(k,i)=(B2(k+1)-mu2(i)*hh(i))/(sigma2(i)*hh(i));
```

```
    Z2(k,i)=((2*pi)^(-1/2))*exp(-(Y2(k,i)^2)/2);
    N2(k,i)=(sigma2(i)*hh(i))*(((2*pi)^(-1/2))...
        *exp(-(B2(k+1)^2)/2));
    %% rows of D are the elements of the diagonal matrix %%
    D(k,i)=(Z1(k,i)*Z2(k,i))/(N1(k,i)*N2(k,i)) ;
end
%% diagonal matrix=DD %%
DD=diag(D(k,:));
xi(:,a+k)=(A*diag(D(k,:))*xi(:,a+k-1));
ff1=xi(:,a+k);
xhat(:,a+k)=ff1./(sum(ff1));
%% GAMMAS %%
sx=xi(:,a+k-1);
hsum=sum(sx);
for r=1:state
  y1(k,r)=(B1(k+1)-mu1(r)*sx(r))/(sigma1(r)*sx(r));
  z1(k,r)=((2*pi)^(-1/2))*exp(-(y1(k,r)^2)/2);
  n1(k,r)=(sigma1(r)*sx(r))*(((2*pi)^(-1/2))...
      *exp(-(B1(k+1)^2)/2));
  y2(k,r)=(B2(k+1)-mu2(r)*sx(r))/(sigma2(r)*sx(r));
  z2(k,r)=((2*pi)^(-1/2))*exp(-(y2(k,r)^2)/2);
  n2(k,r)=(sigma2(r)*sx(r))*(((2*pi)^(-1/2))...
      *exp(-(B2(k+1)^2)/2));
  GAMMA(k,r)=(z1(k,r)*z2(k,r))/(n1(k,r)*n2(k,r));
  sxi=xi(:,a+k-1);
  hsum=sxi(1)+sxi(2)+sxi(3);
  %% Gamma(J1rX)_k %%
  gammaj1(:,r)=(A*diag(D(k,:))*(gammaj1(:,r))...
      +sxi(r)*GAMMA(k,r)*A(1,r)*E(:,1));
  %% Gamma(J2rX)_k %%
  gammaj2(:,r)=(A*diag(D(k,:))*(gammaj2(:,r))...
      +sxi(r)*GAMMA(k,r)*A(2,r)*E(:,2));
  %% Gamma(J3rX)_k %%
  gammaj3(:,r)=(A*diag(D(k,:))*(gammaj3(:,r))...
      +sxi(r)*GAMMA(k,r)*A(3,r)*E(:,3));
  %% Gamma(OrX)_k %%
  gammao(:,r)=(A*diag(D(k,:))*gammao(:,r)...
      +sxi(r)*GAMMA(k,r)*A*E(:,r));
  %% Gamma(Tr(y)X)_k %%
```

```
        gammatnasdaq(:,r)=(A*diag(D(k,:))*gammatnasdaq(:,r)...
            +sxi(r)*GAMMA(k,r)*B1(k+1)*A*E(:,r));
        gammatdowjones(:,r)=(A*diag(D(k,:))*gammatdowjones(:,r)...
            +sxi(r)*GAMMA(k,r)*B2(k+1)*A*E(:,r));
        %% Gamma(Tr(y^2)X)_k %%
        gammatqnasdaq(:,r)=(A*diag(D(k,:))*gammatqnasdaq(:,r)...
            +sxi(r)*GAMMA(k,r)*B1(k+1)^2*A*E(:,r));
        gammatqdowjones(:,r)=(A*diag(D(k,:))*gammatqdowjones(:,r)...
            +sxi(r)*GAMMA(k,r)*B2(k+1)^2*A*E(:,r));
      end
  end
%%%%%%%%%%%%%%%%%%%%%%%% forecast %%%%%%%%%%%%%%%%%%%%%%%%
%% forecast for NASDAQ %%
for k=1:batch
   chi=xi(:,a+k-2);
   b=sum(chi);
   phat=chi./b ;    %p_hat_k=E[X_k |Y_k]
   d=(A^h-1)*phat;
   sum11=d(1)*exp(mu1(1)+0.5*sigma1(1)^2)+d(2)*exp(mu1(2)+0.5*sigma1(2)^2)...
         +d(3)*exp(mu1(3)+0.5*sigma1(3)^2);
   S1(k+h)=C1(a+k-2)*sum11;
end F1(:,u)=S1(1+h:batch+h)';
%% forecast for DOWJONES %%
for k=1:batch
   chi=xi(:,a+k-2);
   b=sum(chi);
   phat=chi./b ;    %% p_hat_k=E[X_k |Y_k] %%
   d=(A^h)*phat;
   sum21=d(1)*exp(mu2(1)+0.5*sigma2(1)^2)+d(2)*exp(mu2(2)+0.5*sigma1(2)^2)...
         +d(3)*exp(mu2(3)+0.5*sigma2(3)^2);
   S2(k+h)=C2(a+k-2)*sum21;
end F2(:,u)=S2(1+h:batch+h)';
%%%%%%%%%%%%%%%%%%%% parameter updates %%%%%%%%%%%%%%%%%%%%
for i=1:state
   gammaj1z(i)=sum(gammaj1(:,i));
   gammaj2z(i)=sum(gammaj2(:,i));
   gammaj3z(i)=sum(gammaj3(:,i));
   gammaos(i)=sum(gammao(:,i));
   gammatsnasdaq(i)=sum(gammatnasdaq(:,i));
```

```
    gammatsdowjones(i)=sum(gammatdowjones(:,i));
    gammatqsnasdaq(i)=sum(gammatqnasdaq(:,i));
    gammatqsdowjones(i)=sum(gammatqdowjones(:,i));
  end
  for i=1:state
      AA(1,i)=gammaj1z(i)/gammaos(i);
      AA(2,i)=gammaj2z(i)/gammaos(i);
      AA(3,i)=gammaj3z(i)/gammaos(i);
      mu1(i)=gammatsnasdaq(i)/gammaos(i);
      mu2(i)=gammatsdowjones(i)/gammaos(i);
      sigma1(i)=sqrt((gammatqsnasdaq(i)-2*mu1(i)...
          *gammatsnasdaq(i)+mu1(i)^2*gammaos(i))/gammaos(i));
      sigma2(i)=sqrt((gammatqsdowjones(i)-2*mu2(i)...
          *gammatsdowjones(i)+mu2(i)^2*gammaos(i))/gammaos(i));
  end
  A=AA;
  MU11(u,:)=mu1(1);
  MU12(u,:)=mu1(2);
  MU13(u,:)=mu1(3);
  MU21(u,:)=mu2(1);
  MU22(u,:)=mu2(2);
  MU23(u,:)=mu2(3);
  SIGMA11(u,:)=sigma1(1);
  SIGMA12(u,:)=sigma1(2);
  SIGMA13(u,:)=sigma1(3);
  SIGMA21(u,:)=sigma2(1);
  SIGMA22(u,:)=sigma2(2);
  SIGMA23(u,:)=sigma2(3);
  PROB1(u,:)=A(1,1);
  PROB2(u,:)=A(1,2);
  PROB3(u,:)=A(1,3);
  PROB4(u,:)=A(2,1);
  PROB5(u,:)=A(2,2);
  PROB6(u,:)=A(2,3);
  PROB7(u,:)=A(3,1);
  PROB8(u,:)=A(3,2);
  PROB9(u,:)=A(3,3);
  %% next interval %%
  a=a+batch;
```

```
e=e+batch;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% plots %%%%%%%%%%%%%%%%%%
forecast11(:,h)=F1(:);
forecast21(:,h)=F2(:);
m=length(forecast21(:,h));
t=h+1;
forecastNasdaq_day97_07(1:t-1,h)=zeros(t-1,1);
forecastNasdaq_day97_07(t:m,h)=forecast11(1:m-h,h);
K=(1:m);
datap1=NASDAQ9707(1:m);
forecastDowjones_day97_07(1:t-1,h)=zeros(t-1,1);
forecastDowjones_day97_07(t:m,h)=forecast21(1:m-h,h);
datap2=DOW_JONES9707(1:m);
%%save forecast data series%%
save forecastNasdaq_day97_07 forecastNasdaq_day97_07
save forecastDowjones_day97_07 forecastDowjones_day97_07
subplot(3,2,1)
plot(K, datap1, K,forecastNasdaq_day97_07(:,h),...
    K, datap2, K,forecastDowjones_day97_07(:,h))
title('One-step ahead forecast for indeces (batch=10)')
leng=length(MU11);
len=[1:1:leng]';
subplot(3,2,2)
plot(len, MU11, len, MU12, len, MU13)
title('Estimates for mu (NASDAQ)')
subplot(3,2,3)
plot(len, MU21, len, MU22, len, MU23)
title('Estimates for mu (Dow Jones)')
subplot(3,2,4)
plot(len, SIGMA11, len, SIGMA12, len, SIGMA13)
title('Estimates for sigma (NASDAQ)')
subplot(3,2,5)
plot(len, SIGMA21, len, SIGMA22, len, SIGMA23)
title('Estimates for sigma (Dow Jones)')
subplot(3,2,6)
plot(len, PROB1, len, PROB2, len, PROB3, len, PROB4, ...
    len, PROB5, len, PROB6, len, PROB7, len, PROB8, len, PROB9)
title('Estimates for transition probability')
```

```
end
```

E.1.2 Matlab code for the investment strategies

```
%% Investment strategies with HMMs %%
clear;
%% load workspace with NASDAQ and DOWJONES data %%
%% daily closing data for ten years (1997 - 2006), 2516 data points in %%
%% each data series%%
load Nasdaqweekriskadj
load Dowjonesweekriskadj
%% load calculated weights %%
load w1
load w2
%% convert financial time series in matrix %%
%% column names of DJweek: %%
%% 'dates' 'CloseDJ''DJforecast' 'DJreturns' 'DJforecastedreturn' %%
%% 'DJriskadjforecast' %%
%% column names of Nasdaqweek: %%
%% 'dates''NasdaqClose''Nasdaqforecast' 'Nasdaqreturns' %%
%% 'Nasdaqforecastedreturn''Nasdaqriskadjforecast' %%
Nasdaqweekriskadjmatrix = fts2mtx(Nasdaqweekriskadj, 1);
Dowjonesweekriskadjmatrix = fts2mtx(Dowjonesweekriskadj, 1);
%% investment backtest for quarterly data %%
%% 40 quarters in the data set %%
%% interval = 13 weeks + starting point %%
counter=0;
interval=14;
for quart=1:41
 %% actual Nasdaq data and dates of current quarter %%
 C1=Nasdaqweekriskadjmatrix(interval:interval+12,4);
 quartdates1(:,quart)=Nasdaqweekriskadjmatrix(interval:interval+12,1);
 Nasdaqdata(:,quart)=C1;
 %% actual Dow Jones data and dates of current quarter %%
 C2=Dowjonesweekriskadjmatrix(interval:interval+12,4);
 quartdates2(:,quart)=Dowjonesweekriskadjmatrix(interval:interval+12,1);
```

```
 Dowjonesdata(:,quart)=C2;
 %% set date vector %%
 Date(:,quart)=Nasdaqweekriskadjmatrix(interval-1:interval+12,1);
 %% pure investment in Nasdaq or Dow Jones %%
 %% initial budget %%
 BudgetNasdaqweek(quart,1)=100;
 BudgetDJweek(quart,1)=100;
 lenweek=length(C1);
 for i=1:lenweek
   BudgetNasdaqweek(quart,i+1)=BudgetNasdaqweek(quart,i)*(exp(C1(i)));
   BudgetDJweek(quart,i+1)=BudgetDJweek(quart,i)*(exp(C2(i)));
 end
 %% pure investment in either Nasdaq or Dow Jones depending on signal %%
 %% from risk-adjusted forecasted return %%
 %% initial budget %%
 Budget2week(quart,1)=100;
 ForecastRiskadjNasdaqWeek=Nasdaqweekriskadjmatrix(interval:interval+12,6);
 ForecastRiskadjDowjonesWeek=Dowjonesweekriskadjmatrix(interval:interval+12,6);
 lenweek2=length(ForecastRiskadjNasdaqWeek);
 disp('Investmentstrategy with weekly returns')
 disp(' ')
 for i=1:lenweek2
  if ForecastRiskadjNasdaqWeek(i) > ForecastRiskadjDowjonesWeek(i)
   disp('invest in Nasdaq at time')
   disp(i)
   Budget2week(quart,i+1)=Budget2week(quart,i)*(exp(C1(i)));
  else
   disp('invest in Dowjones at time')
   disp(i)
   Budget2week(quart,i+1)=Budget2week(quart,i)*(exp(C2(i)));
  end
 end
 %% comparison of last Invstment step %%
 if (Budget2week(quart,lenweek2)>BudgetNasdaqweek(quart,lenweek2)) &...
         (Budget2week(quart,lenweek2)>BudgetDJweek(quart,lenweek2))
  counter=counter+1
 else
  counter=counter
 end
```

```
%% investment strategy with optimal calculated weights from utility %%
%% function %%
%% initial budget %%
 Budget3week(quart,1)=100;
 disp('Mixed Investmentstrategy with weekly returns')
 disp(' ')
 for i=1:lenweek2
  weight1=w1(interval:interval+12);
  weight2=w2(interval:interval+12);
  disp('invest in Nasdaq');
  disp(weight1(i));
  disp('invest in Dowjones');
  disp(weight2(i));
  Budget3week(quart,i+1)=Budget3week(quart,i)*(exp(weight1(i)*C1(i)+...
      weight2(i)*C2(i)));
end
 %% convert prices to returns %%
 Nasdaqret1(quart,:)=price2ret(BudgetNasdaqweek(quart,:));
 DJret1(quart,:)=price2ret(BudgetDJweek(quart,:));
 strategyret1(quart,:)=price2ret(Budget2week(quart,:));
 mixedstrategyret(quart,:)=price2ret(Budget3week(quart,:));
 %% increase interval counter %%
 interval=interval+13;
 %% calculate functions X_Nasdaq and X_DJ for the current quarter %%
 XNasdaq(quart)=log(Budget2week(quart,14)/100)-...
    log(BudgetNasdaqweek(quart,14)/100);
 XDowJones(quart)=log(Budget2week(quart,14)/100)-...
    log(BudgetDJweek(quart,14)/100);
 XNasdaq2(quart)=log(Budget3week(quart,14)/100)-...
    log(BudgetNasdaqweek(quart,14)/100);
 XDowJones2(quart)=log(Budget3week(quart,14)/100)-...
    log(BudgetDJweek(quart,14)/100);
%% calculate mean and variance for returns of investment strategies %%
MeanNasdaq(quart)=mean(Nasdaqret1(quart,:));
MeanDJ(quart)=mean(DJret1(quart,:));
Meanswitching(quart)=mean(strategyret1(quart,:));
Meanmixedstrategy(quart)=mean(mixedstrategyret(quart,:));
VarNasdaq(quart)=var(Nasdaqret1(quart,:));
VarDJ(quart)=var(DJret1(quart,:));
```

```
  VarSwitchingStr(quart)=var(strategyret1(quart,:));
  VarMixedStr(quart)=var(mixedstrategyret(quart,:));
 end
 MeanReturns=[MeanNasdaq' MeanDJ' Meanswitching' Meanmixedstrategy']
 VarianceReturns=[VarNasdaq' VarDJ' VarSwitchingStr' VarMixedStr']
 %% mean and std of outperformance from mixed strategy compared to Nasdaq %%
 MeanXNasdaq=mean(XNasdaq);
 STDXNasdaq=std(XNasdaq);
 %% mean and std of outperformance from switching strategy compared to DJ %%
 MeanXDowjones=mean(XDowJones);
 STDXDowjones=std(XDowJones);
 %% mean and std of worst performance from switching strategy compared to %%
 %% Nasdaq or DJ %%
 MeanMin=mean(min(XNasdaq,XDowJones));
 STDMin=std(min(XNasdaq,XDowJones));
 %% mean and std of outperformance from mixed strategy compared to Nasdaq %%
 MeanXNasdaq2=mean(XNasdaq2);
 STDXNasdaq2=std(XNasdaq2);
 %% mean and std of outperformance from mixed strategy compared to DJ %%
 MeanXDowjones2=mean(XDowJones2);
 STDXDowjones2=std(XDowJones2);
 X1=[1:1:lenweek2+1]';
 %% plots of investments in each quarter (here quarters 1-12) %%
 for p=1:12
  subplot(4,3,p)
  X1=(Date(:,p));
  plot(X1,BudgetNasdaqweek(p,:),X1, BudgetDJweek(p,:),'-.b',...
      X1,Budget2week(p,:),'-r.',X1,Budget3week(p,:),':r.')
  axis([X1(1) X1(14) 50 150])
  dateaxis('x',1, DateString1)
  axis 'auto y'
 end
```

Appendix F

Appendix for Chapter 8

F.1 MATLAB source code: Implementation of the numerical scheme for the default probabilities and calculations of the swap rates

The default probabilities are computed using the source code in F.1.2 in conjunction with the function defined in F.1.1. The last part of this appendix provides the source code for the calculation of the swap rates based on the derived default probabilities.

F.1.1 MATLAB code of function for the Crank-Nicolson scheme

```
-----------------------------------------------

%% default probabilities with Crank-Nicholson method %%
function
[prob1,prob2]=defprob(Loss,V0,...
    Vmin,r1,r2,T,dt,sigma1,sigma2,Vmax1,Vmax2,dV,lambda)
%%set up grid and adjust increments%%
Vmax=Vmax1;
M=round((Vmax)/dV);
```

```
dV=(Vmax)/M;
N=round(T/dt);
dt=T/N;
%% set up transition matrix Q %%
Q=[lambda lambda; lambda lambda];
%% set up borders %%
vetj=0:M;
V1(:,N+1)=linspace(Vmin, Vmax1, M+1)';
V2(:,N+1)=linspace(Vmin, Vmax2, M+1)';
V(:,N+1)=[V1(:,N+1); V2(:,N+1)];
%% calculate indicator function %%
%% set up matrix for calculation and vectors for time and prob %%
matr=zeros(2*(M+1),N+1);
veti=0:N;
vetj=0:M;
aux1=ones(M);
%% set up boundary conditions %%
for j=1:M+1
 if V(j,N+1)<=Loss
  matr(j,N+1)=1;
 else
  matr(j,N+1)=0;
 end
end
for j=(M+2):2*(M+1)
 if V(j,N+1)<=Loss
  matr(j,N+1)=1;
 else
  matr(j,N+1)=0;
 end
end
matr(1,:)=1;
matr(M+1,:)=0;
matr(M+2,:)=1;
matr(2*(M+1),:)=0;
%% set up the coefficients for the matrix %%
a1=0.25*r1*vetj*dt-0.25*sigma1^2*vetj.^2*dt;
b1=-0.5*sigma1^2*vetj.^2*dt-0.5*Q(1,1)*dt-1;
c1=0.25*r1*vetj*dt+0.25*sigma1^2*vetj.^2*dt;
```

```
d1=+0.5*Q(2,2)*dt*aux1;
a2=0.25*r2*vetj*dt-0.25*sigma2^2*vetj.^2*dt;
b2=-0.5*sigma2^2*vetj.^2*dt-0.5*Q(2,2)*dt-1;
c2=0.25*r2*vetj*dt+0.25*sigma2^2*vetj.^2*dt;
d2=+0.5*Q(1,1)*dt*aux1;
b1rhs=0.5*sigma1^2*vetj.^2*dt+0.5*Q(1,1)*dt-1;
b2rhs=0.5*sigma2^2*vetj.^2*dt+0.5*Q(2,2)*dt-1;
%% calculate part matrices %%
DIAGONAL1=[1 b1(2:M) 1];
DIAGONAL2=[-a1(2:M) 0];
DIAGONAL3=[0 c1(2:M)];
DIAGONAL4=[0 d1(1:M-1) 0];
DIAGONAL5=[0 d2(1:M-1) 0];
DIAGONAL6=[1 b2(2:M) 1];
DIAGONAL7=[-a2(2:M) 0];
DIAGONAL8=[0 c2(2:M)];
%% building the matrix %%
matrixA=diag(DIAGONAL2,-1)+diag(DIAGONAL1)+diag(DIAGONAL3,1);
matrixB=diag(DIAGONAL4);
matrixC=diag(DIAGONAL5);
matrixD=diag(DIAGONAL7,-1)+diag(DIAGONAL6)+diag(DIAGONAL8,1);
MATRIX1=[matrixA matrixB; matrixC matrixD];
%% LU-decomposition %%
[L,U]=lu(MATRIX1);
%% right-hand side matrix %%
M2DIAGONAL1=[1 b1rhs(2:M) 0];
M2DIAGONAL2=[a1(2:M) 0];
M2DIAGONAL3=[0 -c1(2:M)];
M2DIAGONAL4=[0 -d1(1:M-1) 0];
M2DIAGONAL5=[0 -d2(1:M-1) 0];
M2DIAGONAL6=[1 b2rhs(2:M) 0];
M2DIAGONAL7=[a2(2:M) 0];
M2DIAGONAL8=[0 -c2(2:M)];
matrix2A=diag(M2DIAGONAL2,-1)+diag(M2DIAGONAL1)+diag(M2DIAGONAL3,1);
matrix2B=diag(M2DIAGONAL4);
matrix2C=diag(M2DIAGONAL5);
matrix2D=diag(M2DIAGONAL7,-1)+diag(M2DIAGONAL6)+diag(M2DIAGONAL8,1);
MATRIX2=[matrix2A matrix2B; matrix2C matrix2D];
%% solve linear systems %%
```

```
for i=N:-1:1
   matr(1:(2*(M+1)),i)=U\ (L \ (MATRIX2*matr(1:(2*(M+1)),i+1)));
end
%% find closest point to V0 on the grid and return prob %%
down=floor((V0)/dV);
up=ceil((V0)/dV);
if down == up
 defaultprob1=matr(down+1,1);
else
 defaultprob1=matr(down+1,1)+...
 (V0-down*dV)*(matr(up+1,1)-matr(down+1,1))/dV;
end
if down == up
 defaultprob2=matr((M+1)+down+1,1);
else
 defaultprob2=matr((M+1)+down+1,1)+...
 (V0-down*dV)*(matr((M+1)+up+1,1)-matr((M+1)+down+1,1))/dV;
end
%% return default probabilities for case 1, starting in 'good' state and %%
%% case 2, starting in 'bad' state %%
 prob1=defaultprob1;
 prob2=defaultprob2;
```

F.1.2 MATLAB code for the calculation of default probabilities for various barrier levels and different values λ of the generator matrix Q

```
%% Default probabilities for different maturities %%
%% calculated under the real-world measure %%
clear;
load interestrates2state
%% set different loss barriers %%
Loss = [110;120; 130; 140];
%% min of firm's value %%
```

```
Vmin=1;
%% starting point of frim's value %%
V0=300;
%% maximum of firm's value in 'good' or 'bad' state %%
Vmax1=800;
Vmax2=550;
%% set values for rates, either constant or using 'input' %%
rr1=0.1      % input('r1    ')
rr2=0.06     % input('r2:    ')
sigma1=0.2;  % input('sigma1  ')
sigma2=0.4;  % input('sigma2  ')
%% setting grid %%
dt=1/50;
dV=10;
N=12;
%% variables for 1-state setting %%
Vmax=Vmax1;
r=rr1
sigma=sigma1;
%%lambda's for generator mathrix Q %%
lambda1=0.1;
lambda2=0.3;
lambda3=0.5;
lambda4=0.7;
lambda5=0.9;
time1=cputime;
%% calculating default probabilities in regime-switching case for different %%
%% values of the loss barrier and different lambdas, utilizing function %%
%% 'defprob', where 'MCI' in variable name denotes starting in 'good' state %%
%% and 'MCII' starting in bad state %%
for T=1:N
 for i=1:4
  [defprobMCIlambda1(i,T),defprobMCIIlambda1(i,T)]=defprob(Loss(i), V0, ...
    Vmin, rr1, rr2, T, dt, sigma1, sigma2, Vmax1, Vmax2, dV,lambda1);
  [defprobMCIlambda2(i,T),defprobMCIIlambda2(i,T)]=defprob(Loss(i), V0, ...
    Vmin, rr1, rr2, T, dt, sigma1, sigma2, Vmax1, Vmax2, dV,lambda2);
  [defprobMCIlambda3(i,T),defprobMCIIlambda3(i,T)]=defprob(Loss(i), V0, ...
    Vmin, rr1, rr2, T, dt, sigma1, sigma2, Vmax1, Vmax2, dV,lambda3);
  [defprobMCIlambda4(i,T),defprobMCIIlambda4(i,T)]=defprob(Loss(i), V0, ...
```

```
    Vmin, rr1, rr2, T, dt, sigma1, sigma2, Vmax1, Vmax2, dV,lambda4);
  [defprobMCIlambda5(i,T),defprobMCIIlambda5(i,T)]=defprob(Loss(i), V0, ...
    Vmin, rr1, rr2, T, dt, sigma1, sigma2, Vmax1, Vmax2, dV,lambda5);
 end
end
time2=cputime-time1
%% calculating default probabilities in non-regime-switching case %%
for T=1:N
 for i=1:4
  defprob1ST(T)=defprob1state210507(Loss(i), V0, ...
  Vmin, r, T, dt, sigma1, Vmax, dV);
 end
end
```

F.1.3 MATLAB code for calculating the swap rates with derived default probabilities

---

```
%% Swap rates for 1- and 2-state model %%
clear;
%% load defaultprobabilities for regime-switching case %%
%% loaded matrices contain calculated default probabilities for different %%
%% maturities and in this case for different values of sigma 1 from %%
%% sensitivity analysis %%
load sens_sigma1_state1_rn
load sens_sigma1_state2_rn
%% load defaultprobabilities for non-regime-switching case for different %%
%% values of mu and sigma %%
load sens_mu1_rn
load sens_sigma1_rn
%% set parameters %%
mu=[0.01 0.02 0.03 0.04 0.05;
    0.01 0.02 0.03 0.04 0.05];
sigma=[0.1 0.2 0.3 0.4 0.5;
    0.1 0.2 0.3 0.4 0.5];
```

```
%% swap rate for 1-state model %%
```

```
%% considered time period %%
n=10;
for count=1:5
 s1S=[];
 s2S=[];
 s3S=[];
 rr1=mu(1,5);
 %% calculate swap-rate for each maturity %%
 for time=1:n
  r=[rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1]';
 %% set default probabilities used for this maturity %%
  for i=1:time+1
   ProbS(i)=sens_sigma1_rn(i,count);
  end
 %% constant rebate of underlying reference asset %%
  R=0.1;
 %% sum for payoff P %%
  for i=2:time+1
   s1S(i)=(ProbS(i)-ProbS(i-1))*exp(-r(i)*i);
  end
  sum1S=sum(s1S);
  %% payoff of the contingent claim %%
  PS=(1-R)*sum1S;
  %% interval between payments %%
  D=1;
  %% sum of expected present values of fee leg F divided by S %%
  for i=2:time+1
   s2S(i)=D*(1-ProbS(i))*exp(-r(i)*i);
  end
  KS=sum(s2S(i));
  %% sum of expected present values of fee accruals A divided by S %%
  for i=2:time+1
   s3S(i)=D/2*(ProbS(i)-ProbS(i-1))*exp(-r(i)*i);
  end
  MS=sum(s3S(i));
  %% swap rate %%
  Swaprate1stateSSigma1(time,count)=PS/(KS+MS);
end
 len=length(Swaprate1stateSSigma1);
```

```
  X1=1:len;
end save Swaprate1stateSSigma1
Swaprate1stateSSigma1
-------------------------------------------------
%% swap rate for 2state model %%
%% considered time period %%
n=10;
for count=1:5
 sI1=[];
 sII1=[];
 %% set parameters %%
 %% different values for rr1 and rr2 possible for sensitivity of r %%
 rr1=mu(1,5);
 rr2=mu(2,1);
 r1(:,1)=[rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1 rr1]';
 r1(:,2)=[rr2 rr2 rr2 rr2 rr2 rr2 rr2 rr2 rr2 rr2 rr2 rr2]';
 rr12=mu(1,5);
 rr22=mu(2,count);
 r2(:,2)=[rr22 rr22 rr22 rr22 rr22 rr22 rr22 rr22 rr22 rr22 rr22 rr22]';
 r2(:,1)=[rr12 rr12 rr12 rr12 rr12 rr12 rr12 rr12 rr12 rr12 rr12 rr12]';
 %% setting default probability matrices %%
 defprobsens1(:,:)=sens_sigma1_state1_rn
 defprobsens2(:,:)=sens_sigma1_state2_rn
 %% Markov chain X for the model parameters mu, sigma and r %%
 %% initial value %%
 state=2;
 X1=[1 0]';
 X2=[0 1]';
 %% initial value for transition probability matrix %%
 lambda=0.1;
 A=[1-lambda lambda; lambda 1-lambda]
 %% calculate swap-rate for each maturity %%
 for time=1:n
 %% modelling a Markov chain %%
  Xhat1(:,1)=X1;
  Xhat2(:,1)=X2;
  for i=2:time+1
   Xhat1(:,i)=A*Xhat1(:,i-1);
   Xhat2(:,i)=A*Xhat2(:,i-1)
```

```
 end
 %% Setting the default probabilities %%
 for i=1:time+1
  ProbI(i)=defprobsens1(i,count);
  ProbII(i)=defprobsens2(i,count);
 end
 for i=2:time+1
  rinnerproduct1(i)=r1(i,:)*Xhat1(:,i);
  rinnerproduct2(i)=r1(i,:)*Xhat2(:,i);
 end
 %% constant rebate of underlying reference asset %%
 R=0.1;
 %% sum for payoff P %%
 for i=2:time+1
  sI1(i)=(ProbI(i)-ProbI(i-1))*exp(-rinnerproduct1(i)*i);
  sII1(i)=(ProbII(i)-ProbII(i-1))*exp(-rinnerproduct2(i)*i);
 end
 sumI1=sum(sI1);
 sumII1=sum(sII1);
 %% payoff of the contingent claim %%
 PI=(1-R)*sumI1;
 PII=(1-R)*sumII1;
 D=1;
 %% sum of expected present values of fee leg F divided by S %%
 for i=2:time+1
  sI2(i)=D*(1-ProbI(i))*exp(-rinnerproduct1(i)*i);
  sII2(i)=D*(1-ProbII(i))*exp(-rinnerproduct2(i)*i);
 end
 KI=sum(sI2(i));
 KII=sum(sII2(i));
 %% sum of expected present values of fee accruals A divided by S %%
 for i=2:time+1
  sI3(i)=D/2*(ProbI(i)-ProbI(i-1))*exp(-rinnerproduct1(i)*i);
  sII3(i)=D/2*(ProbII(i)-ProbII(i-1))*exp(-rinnerproduct2(i)*i);
 end
 MI=sum(sI3(i));
 MII=sum(sII3(i));
 %% swap rate %%
  Swaprate2statesenssigma1I(time,count)=PI/(KI+MI);
```

```
   Swaprate2statesenssigma1II(time,count)=PII/(KII+MII);
  end
end
len=length(Swaprate2statesenssigma1II(:,5));
XX1=1:len;
save Swaprate2statesenssigma1I Swaprate2statesenssigma1I
save Swaprate2statesenssigma1II Swaprate2statesenssigma1II
%% plot swap rates %%
plot(XX1, Swaprate1stateSSigma1, XX1, Swaprate2statesenssigma1I(:,5),...
    XX1, Swaprate2statesenssigma1II(:,5))
```

Bibliography

[1] Hirotugu Akaike. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and B.F. Csaki, editors, *International Symposium on Information Theory, 2 nd*, pages 267–281. Academiai Kiado, Budapest, 1973.

[2] James H Albert and Siddhartha Chib. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics*, 11(1):1–15, 1993.

[3] Fredrik Andersson, Helmut Mausser, Dan Rosen, and Stanislav Uryasev. Credit risk optimization with conditional value-at-risk criterion. *Mathematical Programming*, 89(2):273–291, 2001.

[4] J. Scott Armstrong and Fred Collopy. Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting*, 8:69–80, 1992.

[5] Philippe Artzner and Freddy Delbaen. 'finem lauda' or the risk of swap. *Insurance: Mathematics and Economics*, 9:295–303, 1990.

[6] Philippe Artzner and Freddy Delbaen. Credit risk and prepayment option. *ASTIN Bulletin*, 22:81–96, 1992.

[7] Philippe Artzner and Freddy Delbaen. Default risk and incomplete insurance markets. *Mathematical Finance*, 5:187–195, 1995.

[8] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[9] Ravi Bansal, A. Roland Gallant, Robert Hussey, and George Tauchen. Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics*, 66(1):251–287, 1995.

[10] Ravi Bansal and Hao Zhou. Term structure of interest rates with regime shifts. *Journal of Finance*, 57(5):1997–2043, 2002.

[11] Leonard E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, 3:1–8, 1972.

[12] Leonard E. Baum and J.A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology. *Bulletin of the American Mathematical Society*, 73:360–363, 1967.

[13] Leonard E. Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, 37:1554–1563, 1966.

[14] Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41:164–171, 1970.

[15] Leonard E. Baum and George R. Sell. Growth functions for functions on manifolds. *Pacific Journal of Mathematics*, 27:211–227, 1968.

[16] Fred E. Benth, Lars Ekeland, Ragnar Hauge, and Bjoern R.F. Nielsen. A note on arbitrage-free pricing of forward contracts in energy markets. *Applied Mathematical Finance*, 10(4):325–336, 2003.

[17] Fred E. Benth and Steen Koekebakker. Stochastic modeling of financial electricity contracts. *Department of Mathematics, University of Oslo, E-print No. 24*, 2005.

[18] Manuele Bicego, Vittorio Murino, and Mário A. T. Figueiredo. A sequential pruning strategy for the selection of the number of states in hidden Markov models. *Pattern Recognition Letters*, 24(9-10):1395–1407, 2003.

[19] Patrick Billingsley. *Probability and measure.* Wiley, New York, 1995.

[20] Fischer Black and John C. Cox. Valuing corporate securities: some effects of bond indenture provisions. *Journal of Finance*, 31:351–367, 1976.

[21] Fischer Black and Piotr Karasinski. Bond and option pricing when short rates are log normal. *Financial Analysts Journal*, 47:52–59, 1991.

[22] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

[23] Nicolas P.B. Bollen, Stephen F. Gray, and Robert E. Whaley. Regime switching in foreign exchange rates: Evidence from currency option prices. *Journal of Econometrics*, 94:239–276, 2000.

[24] Tim Bollerslev. Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[25] Michael J. Brennan. The price of convenience and the valuation of commodity contingent claims. In D. Lund and B. Øksendal, editors, *Stochastic Models and Option Models*, pages 33–71. Elsevier/North-Holland, 1991.

[26] Michael J. Brennan and Eduardo S. Schwartz. Evaluating natural resource investments. *Journal of Business*, 58(2):135–157, 1985.

[27] Damiano Brigo and Fabio Mercurio. *Interest rate models: Theory and practice*. Springer Finance. Springer-Verlag Ltd., Heidelberg, 2001.

[28] John Buffington and Robert J. Elliott. American options with regime switching. *International Journal of Theoretical and Applied Finance*, 5:497–514, 2002.

[29] Olivier Cappé, Eric Moulines, and Tobias Rydén. *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer, New York, 2005.

[30] Rene Carmona and Michael Ludkovski. Spot convenience yield models for the energy markets. In G. Yin and Q. Zhang, editors, *Mathematics of Finance*, volume 351 of *Contemporary Mathematics*, pages 65–79. American Mathematical Society, 2004.

[31] Álvaro Cartea and Marcelo G. Figueroa. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335, 2005.

[32] Hanjie Chen and John N. Jiang. Empirical electricity price modelling. *Energy Risk*, July:74–79, 2006.

[33] Michael P. Clements, Philip Hans Franses, and Norman R. Swanson. Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, 20(2):169–183, 2004.

[34] Les Clewlow and Chris Strickland. Valuing energy options in a one factor model fitted to forward prices. *Quantitative Finance Research Group, Working paper, University of Technology, Sydney, Australia*, 1999.

[35] John C Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):363–384, 1985.

[36] Michel Culot, Valérie Goffin, Steve Lawford, Sébastien de Menten, and Yves Smeers. An affine jump diffusion model for electricity. Technical report, Department of Mathematical Engineering and CORE, Catholic University of Leuven, 2006.

[37] Qiang Dai and Kenneth Singleton. Term structure dynamics in theory and reality. *Review of Financial Studies*, 16:631–678, 2003.

[38] Cyriel de Jong. The nature of power spikes: a regime-switch approach. *Studies in Nonlinear Dynamics & Econometrics*, 10, 2006. http://www.bepress.com/snde/vol10/iss3/art3.

[39] Cyriel de Jong and Ronald Huisman. Option formulas for mean-reverting power prices with spikes. Technical report, Erasmus Research Institute of Management,Erasmus University Rotterdam, 2002.

[40] Angus Deaton and Guy Larouque. Competitive storage and commodity price dynamics. *Journal of Political Economy*, 104:896–923, 1996.

[41] Freddy Delbaen and Walter Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300:463–520, 1994.

[42] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B. Methodological*, 39(1):1–38, 1977.

[43] Shijie Deng. Stochastic models of energy commodity prices and their applications: mean-reversion with jumps and spikes. Technical report, University of California Energy Institute, 2000.

[44] Francis X. Diebold and Lutz Kilian. Measuring predictability: theory and macroeconomic applications. *Journal of Applied Econometrics*, 16:657–669, 2001.

[45] John Driffill, Turalay Kenc, and Martin Sola. An empirical examination of term structure models with regime shifts. Technical report, Birkbeck University of London, 2003.

[46] Darrell Duffie and Raymond Kan. A yield-factor model of interest rates. *Mathematical Finance*, 64:379–406, 1996.

[47] Darrell Duffie and Kenneth J. Singleton. Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12:687–720, 1999.

[48] Robert J. Elliott. Exact adaptive filters for Markov chains observed in Gaussian noise. *Automatica*, 30(9):1399–1408, 1994.

[49] Robert J. Elliott, Lakhdar Aggoun, and John B. Moore. *Hidden Markov models*, volume 29 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1995.

[50] Robert J. Elliott, Leunglung Chan, and Tak Kuen Siu. Option pricing and Esscher transform under regime switching. *Annals of Finance*, 1:423–432, 2005.

[51] Robert J. Elliott, Paul Fischer, and Eckhard Platen. Filtering and parameter estimation for a mean reverting interest rate model. *Canadian Applied Mathematics Quarterly*, 7:381–400, 1999.

[52] Robert J. Elliott and Juri Hinz. A method for portfolio choice. *Applied Stochastic Models in Business and Industry*, 19:1–11, 2003.

[53] Robert J. Elliott, William C. Hunter, and Barbara M. Jamieson. Financial signal processing: a self calibrating model. *International Journal of Theoretical and Applied Finance*, 4(4):567–584, 2001.

[54] Robert J. Elliott and Ekkehard P. Kopp. *Mathematics of financial markets*, volume 2nd ed. Springer, New York, 2005.

[55] Robert J. Elliott and Vikram Krishnamurthy. New finite-dimensional filters for parameter estimation of discrete-time linear gaussian models. *IEEE Transactions on Automatic Control*, 44(5):938–951, 1999.

[56] Robert J. Elliott and Rogemar S. Mamon. An interst rate model with a Markovian mean reverting level. *Quantitative Finance*, 2(6):454–458, 2002.

[57] Robert J. Elliott, Gordon A. Sick, and Michael Stein. Modelling electricity price risk. *Preprint, University of Calgary*, 2003.

[58] Robert J. Elliott and John van der Hoek. An application of hidden Markov models to asset allocation problems. *Finance and Stochastics*, 1:229–238, 1997.

[59] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[60] Yariv Ephraim and Neri Merhav. Hidden Markov processes. *IEEE Transactions on Information Theory*, 48(6):1518–1569, 2002.

[61] Christina Erlwein and Rogemar Mamon. An online estimation scheme for a Hull-White model with HMM-driven parameters. *Statistical Methods and Applications*, DOI: 10.1007/s10260-007-0082-4, available online: http://www.springerlink.com/content/6457868m38527757, 2007.

[62] Christina Erlwein, Rogemar Mamon, and Fred E. Benth. HMM filtering and parameter estimation of an electricity spot price model. Technical report, Department of Mathematics, University of Oslo, E-print No. 2, 2007.

[63] Christina Erlwein, Gautam Mitra, and Diana Roman. HMM based scenario generation for an investment optimization problem. Technical report, CTR/68/07, CARISMA, Brunel University, 2007.

[64] Martin D. Evans. Real risk, inflation risk and the term structure. *Economic Journal*, 113:345–389, 2003.

[65] Michael J. Evans and Jeffrey S. Rosenthal. *Probability and statistics: the science of uncertainty.* WH Freeman and Company, New York, 2004.

[66] Natalia Fabra and Juan Toro. Price wars and collusion in the Spanish electricity market. *International Journal of Industrial Organization*, 23(3-4):155–181, 2005.

[67] Eugene F. Fama and Michael R. Gibbons. A comparison of inflation forecasts. *Journal of Monetary Economics*, 13:327–348, 1984.

[68] Rene Garcia and Pierre Perron. An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics*, 78(1):111–125, 1996.

[69] Paul H. Garthwaite, Ian T. Jolliffe, and Byron Jones. *Statistical inference.* Oxford University Press, 2002.

[70] Hélyette Geman and Andrea Roncoroni. Understanding the fine structure of electricity prices. *Journal of Business*, 79(3), 2006.

[71] Hans U. Gerber and Elias S.W. Shiu. Option pricing by Esscher transforms (with discussions). *Transactions of the Society of Actuaries*, 46:99–191, 1994.

[72] Hans U. Gerber and Elias S.W. Shiu. Discussion on valuing equity-indexed annuitues by Serena Tiong. *North American Actuarial Journal*, 4(4):164–169, 2000.

[73] Rajna Gibson and Eduardo S. Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, 45(3):959–976, 1990.

[74] Alicia M. González, Antonio M. San Roque, and Javier García-González. Modeling and forecasting electricity prices with input/output hidden Markov models. *IEEE Transactions on Power Systems*, 20(1):13, 2005.

[75] Stephen F. Gray. Modelling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62, 1996.

[76] Xin Guo. Information and option pricings. *Quantitative Finance*, 1:38–44, 2001.

[77] Dirk Hackbarth, Jianjun Miao, and Erwan Morellec. Capital structure, credit risk and macroeconomic conditions. *Journal of Financial Economics*, 82:519–550, 2006.

[78] James D. Hamilton. Rational-expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, 12(2-3):385–423, 1988.

[79] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and business cycle. *Econometrica*, 57(2):357–384, 1989.

[80] James D. Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70, 1990.

[81] James D. Hamilton. *Time series analysis.* Princeton University Press, New Jersey, 1994.

[82] Bruce E. Hansen. The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics*, 7:61–82, 1992.

[83] Bruce E. Hansen. Erratum: the likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics*, 11(2):195–198, 1996.

[84] Floyd B. Hanson and John J. Westman. Stochastic analysis of jump-diffusions for financial log-return processes. *Stochastic Theory and Control, Proceedings of a Workshop held in Lawrence, Kansas, October*, pages 18–20, 2001.

[85] Mary R. Hardy. A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 2:41–53, 2001.

[86] J. Michael Harrison and David M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408, 1979.

[87] J. Michael Harrison and Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications*, 11:215–280, 1981.

[88] J. Michael Harrison and Stanley R. Pliska. A stochastic calculus model of continuous trading: complete markets. *Stochastic Processes and Their Applications*, 15:313–316, 1983.

[89] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: a new methodology. *Econometrica*, 60:77–105, 1992.

[90] Ronald Huisman and Ronald Mahieu. Regime jumps in electricity prices. *Energy Economics*, 25(5):425–434, 2003.

[91] John C. Hull and Alan White. Pricing interest rate derivative securities. *Review of Financial Studies*, 3:573–592, 1990.

[92] Robert Jarrow and Stuart Turnbull. Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, 50:53–85, 1995.

[93] Rüdiger Kiesel, Gero Schindlmayr, and Reik H. Börger. A two-factor model for the electricity forward market. Technical report, Institute of Mathematical Finance, Ulm University, 2007.

[94] Chang-Jin Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1):22, 1994.

[95] Tino Kluge. *Pricing swing options and other electricity derivatives.* PhD thesis, University of Oxford, 2006.

[96] Pavlo Krokhmal, Jonas Palmquist, and Stanislav Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4(2):21–41, 2002.

[97] Camilla Landén. Bond pricing in a hidden Markov model of the short rate. *Finance and Stochastics*, 4(4):371–389, 2000.

[98] David Lando. On Cox processes and credit risky securities. *Review of Derivatives Research*, 2:99–120, 1998.

[99] David Lando. *Credit risk modelling: theory and applications.* Princeton Series in Finance. Princeton University Press, 2004.

[100] Hayne E. Leland. Agency costs, risk management, and capital structure. *Journal of Finance*, 53:1213–1243, 1998.

[101] Hayne E. Leland. Predictions of default probabilities in structural models of debt. *Journal of Investment Management*, 2:1–28, 2004.

[102] Lennart Ljung. *System identification - theory for the user.* Prentice-Hall, 2nd edition, 1999.

[103] Francis Longstaff and Eduardo Schwartz. Interest volatility and the term structure: a two-factor general equilibrium model. *Journal of Finance*, 47:1259–1282, 1992.

[104] Francis Longstaff and Eduardo S. Schwartz. A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance*, 50:789–819, 1995.

[105] Julio J. Lucia and Eduardo S. Schwartz. Electricity prices and power derivatives: evidence from the Nordic power exchange. *Review of Derivatives Research*, 5(1):5–50, 2002.

[106] Shangzhen Luo. Multi-period asset allocation under hidden Markovianly driven noises. *Stochastic Analysis and Applications*, 25(5):1057–1078, 2007.

[107] Helmut Lütkepohl. *New introduction to multiple time series analysis.* Springer, Berlin, 2005.

[108] Dilip B. Madan and Haluk Unal. Pricing the risks of default. *Review of Derivatives Research*, 2:121–160, 1998.

[109] Rogemar S. Mamon. *Market models of interest rate dynamics with a joint short rate/HJM approach.* PhD thesis, University of Alberta, 2000.

[110] Rogemar S. Mamon. A time-varying Markov chain model of term structure. *Statistics and Probability Letters*, 60(3):309–312, 2002.

[111] Rogemar S. Mamon, Christina Erlwein, and R. Bhushan Gopaluni. Adaptive signal processing of asset price dynamics with predictability analysis. *Information Sciences*, 178(1):203–219, 2008.

[112] Harry Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[113] MathWorks Inc. GARCH toolbox user's guide. *MATLAB Version 6.5*, 2002.

[114] Geoffrey J. McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions.* Wiley, New York, 1997.

[115] Robert C. Merton. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29:449–470, 1974.

[116] Robert C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.

[117] Kristian R. Miltersen. Commodity price modelling that matches current observables: a new approach. *Quantitative Finance*, 3:51–58, 2003.

[118] Sam Mirmirami and H.C. Li. Gold price, neural networks and genetic algorithm. *Computational Economics*, 23:193–200, 2004.

[119] John M. Mulvey. Introduction to financial optimization: mathematical programming special issue. *Mathematical Programming*, 89(2):205–216, 2001.

[120] Marek Musiela and Marek Rutkowski. *Martingale methods in financial modelling.* 2nd ed. Springer, New York, 1997.

[121] Vasant Naik and Moon H. Lee. Yield curve dynamics with discrete shifts in economic regimes: theory and estimation. Working paper, University of British Columbia, 1997.

[122] Elisa Nicolato and Emmanouil Venardos. Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. *Mathematical Finance*, 13(4):445–466, 2003.

[123] James R. Norris. *Markov chains.* Springer Finance. Cambridge University Press, Cambridge, 1997.

[124] Madalina Olteanu. A descriptive method to evaluate the number of regimes in a switching autoregressive model. *Neural Networks*, 21(4):963–972, 2006.

[125] Edoardo Otranto and Giampiero M. Gallo. A nonparametric Bayesian approach to detect the number of regimes in Markov switching models. *Econometric Reviews*, 21(4):477–496, 2002.

[126] Antoon Pelsser. *Efficient methods for valuing interest rate derivatives.* Springer Finance. Springer-Verlag London Ltd., London, 2000.

[127] Georg C. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In S. Uryasev, editor, *Probabilistic Constrained Optimization: Methodology and Applications*, volume 38, pages 272–281. Kluwer Academic Publishers, 2000.

[128] Zacharias Psaradakis and Nicola Spagnolo. On the determination of the number of regimes in Markov-switching autoregressive models. *Journal of Time Series Analysis*, 24(2):237–252, 2003.

[129] Lawrence R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.

[130] R. Tyrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

[131] R. Tyrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471, 2002.

[132] Diana Roman, Ken Darby-Dowman, and Gautam Mitra. Portfolio construction based on stochastic dominance and target return distributions. *Mathematical Programming*, 108(2):541–569, 2007.

[133] Tobias Rydén, Timo Teräsvirta, and Stefan Åsbrink. Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13(3):217–244, 1998.

[134] Eduardo S. Schwartz. The stochastic behaviour of commodity prices: implications for valuation and hedging. *Journal of Finance*, 52:923–973, 1997.

[135] Eduardo S. Schwartz. Valuing long-term commodity assets. *Financial Management*, 27:57–66, 1998.

[136] Eduardo S. Schwartz and James E. Smith. Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46:893–911, 2000.

[137] Neil Shephard. Statistical aspects of ARCH and stochastic volatility. In D.R. Cox D.V. Hinkley and O.E. Barndorff-Nielsen, editors, *Time series models in econometrics, finance and other fields*, pages 1–67. Chapman & Hall, London, 1996.

[138] Albert N. Shiryaev. *Essentials of stochastic finance.* World Scientific River Edge, New Jersey, 1999.

[139] Tak Kuen Siu, Christina Erlwein, and Rogemar Mamon. The pricing of credit default swaps under a Markov-modulated Merton's structural model. *North American Actuarial Journal*, forthcoming.

[140] Daniel R. Smith. Markov-switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business and Economic Statistics*, 20(2):183–197, 2002.

[141] Birgit Strikholm and Timo Teräsvirta. A sequential procedure for determining the number of regimes in a threshold autoregressive model. *Econometric Journal*, 9:472–491, 2006.

[142] Nikola A. Tarashev. An empirical evaluation of structural credit risk models. Working paper, Bank for International Settlements, 2005.

[143] Stephen J. Taylor. *Asset price dynamics, volatility and prediction.* Princeton University Press, Princeton, 2005.

[144] Nikolas Topaloglou, Hercules Vladimirou, and Stavros A. Zenios. CVaR models with selective hedging for international asset allocation. *Journal of Banking and Finance*, 26(7):1535–1561, 2002.

[145] Oldrich A. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

[146] Paul Wilmott. *Paul Wilmott on Quantitative Finance.* John Wiley and Sons, New York, 2000.

[147] C. F. Jeff Wu. On the convergence properties of the em algorithm. *Annals of Statistics*, 11:95–103, 1983.

[148] Wang Yu and Gerald B. Sheblé. Modeling electricity markets with hidden Markov model. *Electric Power Systems Research*, 76(6-7):445–451, 2006.

[149] Moshe Zakai. On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 11:230–243, 1969.

[150] Chunsheng Zhou. An analysis of default correlations and multiple defaults. *Review of Financial Studies*, 14:555–576, 2001.

[151] Chunsheng Zhou. The term structure of credit spreads with jump risk. *Journal of Banking and Finance*, 25:2015–2040, 2001.